AD-A167 781

PERFORMANCE OF A WORKING
FACE RECOGNITION MACHINE
USING CORTICAL THOUGHT THEORY

THESIS

Robert L. Russel Jr.
Captain, USAF

AFIT/GE/ENG/85D-37

DEPARTMENT OF THE AIR FORCE
AIR UNIVERSITY
AIR FORCE INSTITUTE OF TECHNOLOGY

Wright-Patterson Air Force Base, Ohio

AFIT/GE/ENG/85D-37

PERFORMANCE OF A WORKING
FACE RECOGNITION MACHINE
USING CORTICAL THOUGHT THEORY

THESIS

Robert L. Russel Jr.
Captain, USAF

AFIT/GE/ENG/85D-37

AFIT/GE/ENG/85D-37

PERFORMANCE OF A WORKING FACE RECOGNITION MACHINE USING CORTICAL THOUGHT THEORY

THESIS

Presented to the Faculty of the School of Engineering

of the Air Force Institute of Technology

Air University

In Partial Fulfillment of the

Requirements for the Degree of

Master of Science in Electrical Engineering

Robert L. Russel Jr., BSEE

Captain, USAF

December 1985

Acknowledgements

In a very real way, this work is a combination of the ideas and efforts of four people, inseparable from any one of them: myself, Dr. Richard Routh, Dr. James Holten, and Dr. Matthew Kabrisky. I consider it one of my greatest privileges to have worked with these superb minds.

The Holten and Routh families provided me with extra special love and support during the Master's program, and have my utmost love and gratitude. Jim and Mona Holten, along with Ric and Edie Routh, are among the finest and most impressive people I've ever met. Mona and Erin Holten did the majority of the artwork in this thesis, and Jim spent a lot of time in order to print the thesis in a quality manner. Without their help, I would never have finished on time.

Special thanks is given to Mr. Dan Zambon for his excellent engineering support in keeping the computers up and running, and for his graciously providing for the special system requirements of the CTT Face Recognition System.

As a Christian, I have seen the hand of God at work in in many ways in providing direction and blessing my research, totally apart from my deserving it. The first and foremost credit and thanks for this work must therefore go to the Lord Jesus Christ, to whom I initially gave ownership and authority over this thesis effort. I hope it is of a quality which will give glory to Him.

Table of Contents

List of Figures

List of Tables

AFIT/GE/ENG/85D

ABSTRACT

A face recognition system was developed, based on the principles of Cortical Thought Theory (CTT), recently proposed by Dr. Richard L. Routh as his doctoral dissertation at the Air Force Institute of Technology. Routh tested the CTT architecture successfully for speech processing. In order to evaluate this architecture as a generic sensory information processing model, CTT was tested for visual processing, specifically for the difficult task of human face recognition.

The CTT gestalt transformation maps a 2-dimensional image into a 2-D coordinate point. The present system extracts six sub-images from a contrast-expanded image, calculates the 2-D gestalt coordinates, and stores the information in a database. Statistics are then calculated on at least five prototypes processed for each person. Overall performance of different sub-windows on a face are also determined. An "unidentified" person is recognized by calculating the six gestalt feature vectors, and then finding the closest match to previously stored data. The computer generates an ordered list by closeness of match. Performance testing of the system yielded a reliability of 90%.

The system exhibits many characteristics of human recognition. The following are the significant results of this research:

1) Provides a possible explanation of why the primate visual system splits images vertically before displaying them on separate right and left primary visual cortexes.

2) Provides a plausible explanation of why humans experience difficulty in recognizing negative images.

3) Faces which look similar to humans map close together in CTT space, and faces which look quite different to humans map far apart in CTT space.

4) Partial face images which seem to give the highest recognition performance in human psychological experiments give the highest performance in the CTT model.

5) The system is reasonably consistent with the human physiology as it is presently understood.

The performance of the face recognition system strongly suggests CTT's general applicability to vision, and increases its credibility as a general model of human sensory information processing.

I. Introduction

This investigation evaluated a new unified brain theory, called Cortical Thought Theory (CTT), in the domain of vision by using the CTT principles to try to build a face recognition system. This research concludes that CTT is indeed applicable to vision, and this document describes the design, implementation, and performance of a working face recognition machine built solidly upon the principles of CTT.

For years, scientists have been enamored with the prospect of designing machines that process information in a manner similar to humans. This has spawned much of the current effort in artificial intelligence. Pattern recognition, the ability to "recognize" something (such as audio or visual inputs), has been one of the more difficult skills to copy. In fact, it has been said that the typical two-year-old can do a better job of pattern recognition than the best of our supercomputer systems (14). In addition, the human brain seems to be able to do nearly instantaneous direct-memory access to the most important piece of information in an adult-size knowledge base (21). Current systems, however, experience an exponential growth in search time as the size of the knowledge base increases. These and other problems have led many researchers to conclude that the processing and architecture of the human brain are fundamentally different than our current computer

architectures.

Dr. Richard L. Routh, in his doctoral work at the Air Force Institute of Technology (1983-1985), developed what he claims to be a general model of human thought processing, dubbed "Cortical Thought Theory." He demonstrated it successfully on a limited scale in speech recognition, and predicted and verified a new class of audio illusions. In order for this work to be accepted as a general model, however, it must go through a series of tests which demonstrates its applicability in various human information processing tasks.

Since the structure of a mechanism implies its function, and the cortex has basically the same structure across its entire surface, then the basic mechanism used to process information in one domain, such as speech, must also apply to vision, higher-level thinking, and all other processes in the cortex. The applicability of CTT to visual processing would therefore be a major test of its applicability as a general model of human information processing. What kind of visual test should be used? A trivial task would prove little. Human face recognition, however, is considered an extremely difficult problem. Successful demonstration of a face recognition machine, built solidly upon CTT principles, would strongly suggest CTT's general applicability to the domain of vision, and increase its credibility as a model of human information

processing.

<u>Criteria</u>. What criteria will be applied to evaluate the evidence?

1) The system must demonstrate "human-like" classification of human face images. For instance, an image that a human would classify as similar must be mapped close together in some discrimination space, while images that look dissimilar to humans must be mapped far apart. (Additional characteristics of a human-like recognition system are listed in the next chapter.)

2) The system must demonstrate repeatability. In this specific application, the system must map an image of a person's face onto the same general coordinate location for each facial image processed, as opposed to mapping an image to one end of the coordinate space for one image of a person, and to the other end for another image of the same person.

3) The system must achieve reasonably high performance by identifying randomly-selected human face images with an accuracy of at least 90% (an arbitrary value.)

4) All critical components and assumptions of the system must be consistent with CTT, and the human physiology to the extent it is presently understood.

5) To satisfy the original challenge of my thesis advisor, Dr. Matthew Kabrisky, the system must be able to distinguish the two subjects in figure 1-1.

a. Capt Fred Stierwalt

b. Capt Tom Griffin

Figure 1-1. Stierwalt vs Griffin--the Original Challenge

Assumptions. The following assumptions were made to reduce the problem to a manageable level:

1) Images of a person's face are captured and digitized in the laboratory, as opposed to being taken from photographs. This means that the subject's position and lighting can be controlled.

2) The background behind the subject is white cardboard, and the lighting is constant.

3) The subject is not moving and not smiling.

4) The subject is looking straight at the camera, with no rotations from a nomimal full face view.

5) Subject variability is not significant from picture to picture (e.g., no radical hairstyle changes.)

6) 5 pictures per subject are adequate to characterize a person.

7) 20 subjects are sufficient to prove the concept.

These assumptions will greatly simplify the problem. Will they, however, oversimplify the problem so that the results are useless? No, as the limitations imposed by the assumptions can most likely be removed in an advanced system by using already-established techniques.

Overview. The background of previous efforts in machine face recognition and the human ability to recognize faces is discussed in chapter 2, along with a summary of proposed characteristics of a human-like face recognition system. Chapter 3 discusses in detail the background of

Cortical Thought Theory. Chapter 4 discusses the design of our face recognition system that is based on the principles of CTT. Chapter 5 describes the system as actually implemented. Chapter 6 deals with testing, results, and limitations. Chapter 7 gives a summary and conclusions, and Chapter 8 gives recommendations for further research.

II. Background of Previous Work in Human Face Recognition

There have been relatively few major attempts at machine face recognition over the years, but there has been a moderate amount of psychological testing on the human's ability to recognize faces. This section will discuss the key studies that have been done in these areas and summarize with a list of characteristics which should be exhibited by a system having "human-like" face recognition qualities.

Bledsoe. The first major work in machine facial recognition was done by Dr. Woodrow W. Bledsoe at Panoramic Research in 1966 (2). The problem he attempted to solve was quite ambitious, involving recognition of a face in a photograph where there may be a great variability in head rotation and tilt, lighting intensity and angle, facial expression, etc. His sample set included about 2000 photographs, with at least two poses (usually exactly two) for each person. Bledsoe reported, "This sample [set] contained examples of every conceivable combination of head rotation, tilt, and lean (within limits), and included a realistic variation photographic quality and light contrasts."

The feature set used was a series of coordinates and their ratios for certain key points on the face (see figure 2-1.) The points were located and entered by a human operator using a digitizing tablet. Among the points

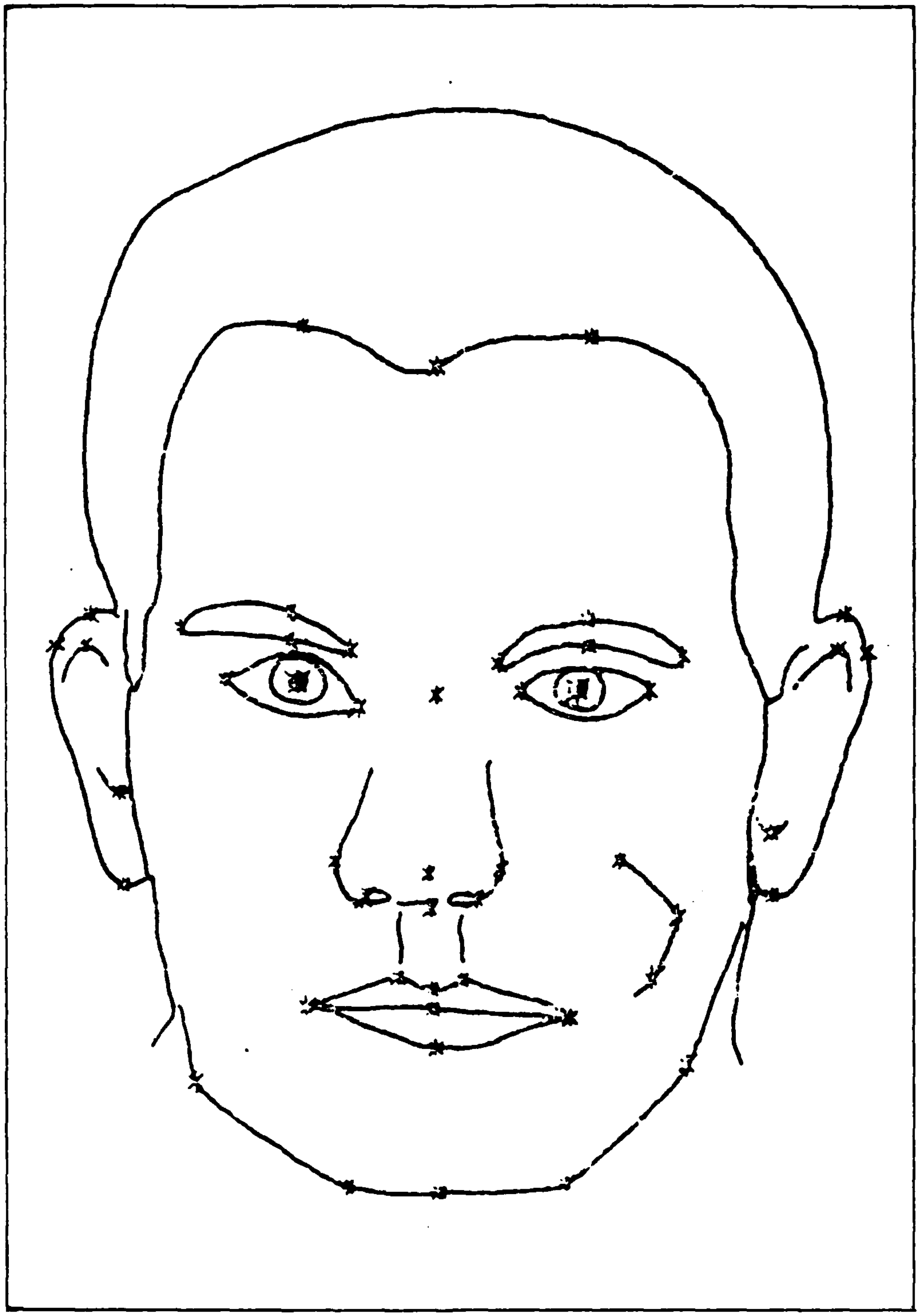


Figure 2-1. Original Measurements made on Face by
Source: (2) Bledsoe

selected were varied locations of key features on the eyes, ears, nose, mouth, chin, eyebrows, and hairline. A complete list is given in table 2-1.

Before a distance measurement was processed, the points were normalized for scale, by dividing all the distances by the distance between the two pupils of the eyes. In addition, the rotation, tilt, and lean of the head were estimated, and then the coordinate points were "rotated" back to a frontal pose. This procedure was reported to have worked quite successfully, with its errors in estimating angles contributing little to the overall error in the system.

Recognition was attempted by comparing pairs of photographs. A "pseudo-distance" was computed between the two photographs, as follows:

$$d^2 = \frac{1}{K} \sum_{i=1}^{K} \frac{|D_{i1} - D_{i2}|^2}{\sigma_{Ei}} \qquad (2\text{-}1)$$

where the D's are the normalized distances between certain fine features of the face, and the σ_{Ei}'s are the standard deviations of measurement errors.

The "goodness" of the system was measured by the average number of incorrect names that were placed ahead of the true name in the identity list provided by the computer. If a total file contained N photographs, and the average number

Name ______________ Date __________ Repeat No. ______ Picture No. __

	Left x	y	Center	Right x	y
1. Eyes					
1. Center of pupils	____	____		____	____
2. Outside corners	____	____		____	____
3. Inside corners	____	____		____	____
2. Ears					
1. Top tips	____	____		____	____
2. Bottom tips	____	____		____	____
3. Inner top	____	____		____	____
4. Inner bottom	____	____		____	____
5. Point opposite 3	____	____		____	____
3. Nose					
1. Highest point on line between eyes		____		____	
2. Tip (highest point)		____		____	
3. Base		____		____	
4. Widest points	____	____		____	____
5. Outside points of nostrils	____	____		____	____
4. Mouth					
1. Center tip of upper lip		____		____	
2. Center edge of lower lip		____		____	
3. Center of mouth		____		____	
4. Upper tips of bottom of vertical ridges	____	____		____	____
5. Outside corners of mouth	____	____		____	____
5. Chin					
1. Point		____		____	____
2. 2 side points	____	____		____	____
3. Corners of jaw	____	____		____	____
6. Eyebrows					
1. Midpoint top	____	____		____	____
2. Midpoint bottom	____	____		____	____
3. Outside end	____	____		____	____
4. Inside end	____	____		____	____
7. Hairline					
1. Point of widows peak		____		____	
2. Intersection of vertical line through eyes	____	____		____	____

LIST OF POINTS ON FACE WHERE COORDINATES ARE OBTAINED

Table 2-1.
Source:(2)

of incorrect names was S, then

$$F=S/N, \tag{2-2}$$

where F was called the Average Reduction in Uncertainty (1). He reported an F of about 0.01 (or less in some cases), meaning that, on the average, the correct name was usually within the top 1% of names in the identity list provided by the computer.

He found a better distance measurement, however, to be the best fit between subsets of one picture and subsets of another. (The absolute location of the subsets on the face were ignored.) For instance, the subset of points containing the eyes in the picture to be recognized is matched to the same subset in all the training pictures to find the best fit (see figure 2-2.)

Bledsoe reported the following as the problems in the system:

> When two photographs of the same person failed to be identified, the reason was usually one or more of the following:
> (1) Poor-quality photograph, in texture, lighting, contrast
> (2) Operator or machine error in giving a coordinate
> (3) Difference in expression such as smile, mouth open, eyebrows raised
> (4) Large differences in angles of head rotation
> Usually one of these was sufficient to cause a large jump in the pseudo-distance, d, and thereby (effectively) prevent recognition.
> However, there were cases which failed for no reason, and it is still not clear why this method did not do even better

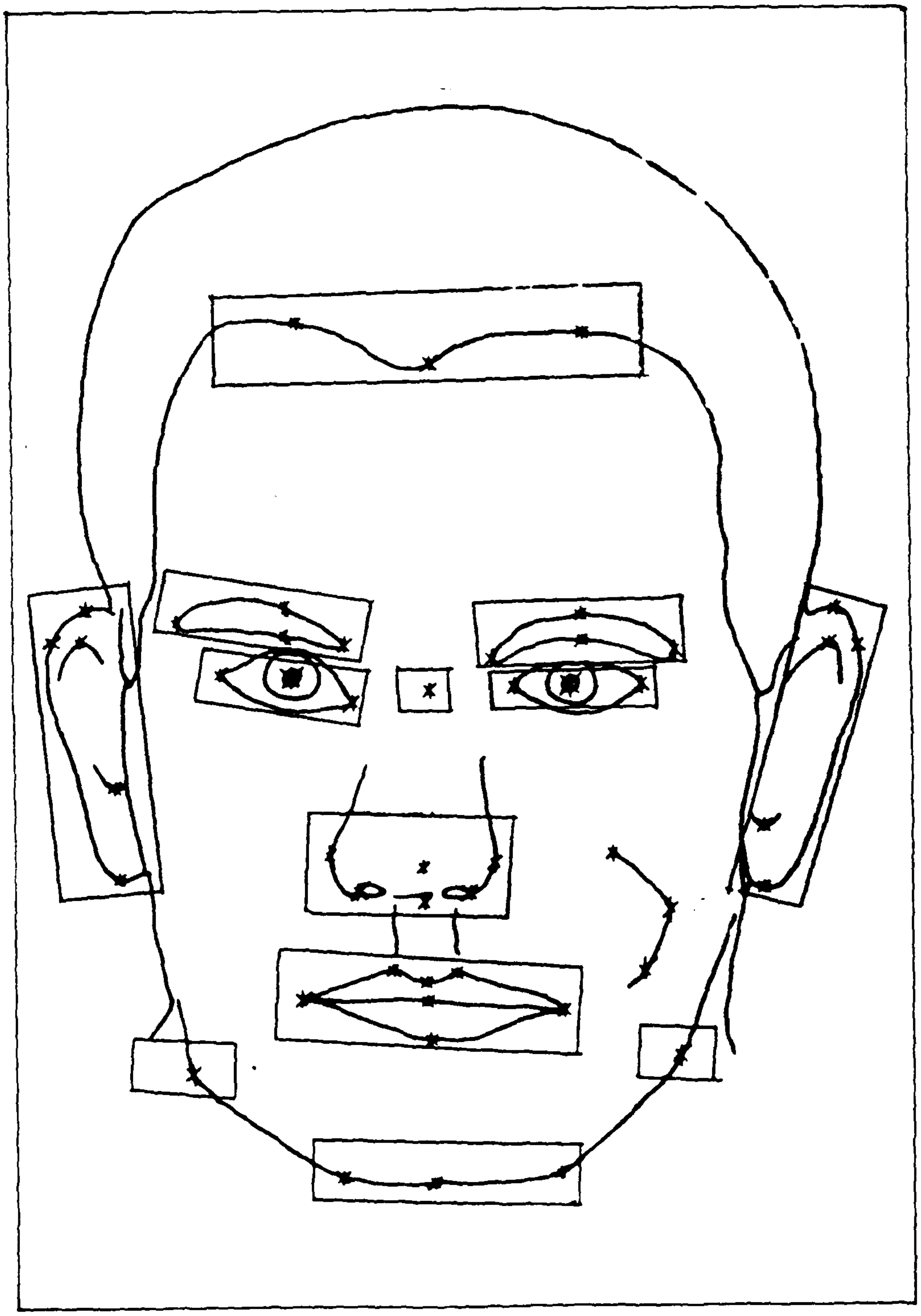


Figure 2-2. Points on the Face Divided into Natural
Source: (2) Subsets

on this large file of photographs.
Undoubtedly, there are pertinent factors
that have not been considered (2).

Recommendations from Bledsoe's study included research on facial-recognition systems that are "completely automatic (remove the man from the man-machine system), and on other systems such as recognition from stereo pairs."

Other experiments were tried which shed some light on the difficulty of face recognition. One was to have the computer locate the key features and determine their coordinates. Bledsoe reported, "Such a technique has not yet been satisfactorily developed because of the difficulty in the step of computer location of facial features, but such a development seems feasible within the near future (2)." (See figure 2-3.) No further work in this area has been reported by Bledsoe.

Harmon. Leon D. Harmon reported work on face recognition in two papers (10, 7), with the second one co-authored with Goldstein and Lesk. In his first paper, he began by citing previous approaches in this area. For instance, he cited work by Fennema and Hart as having, like Bledsoe, a semi-automated approach which sorted and classified rather than identified uniquely (10:196). Harmon also discussed attempts to aid automatic facial feature analysis, including machines which optically generate

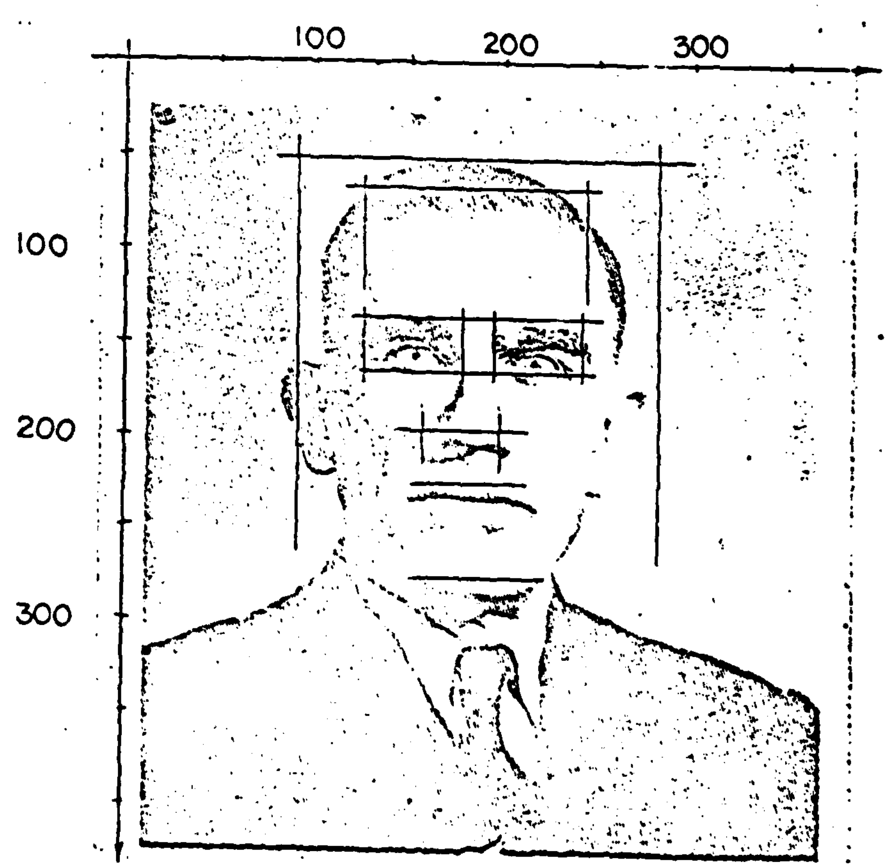


Figure 2-3. Features Located by Machine
Source: (1)

contour-line plots of faces (3:769-782,10:196), and machines which evaluate skull measurements (3:196).

Harmon set out to better examine human capabilities for face recognition, and thereby help to define features useful for machine recognition.

Harmon's approach in his first paper was referred to as "analysis through synthesis." The basic idea of this approach is that if variables can be discovered which will create, or "synthesize", certain patterns in a reproducable manner, then the values of the variables can be used as "features" for detection of these patterns in subsequent recognition (10). First, Harmon tested to see if the descriptor set used in police work would allow "synthesis" of a recognizable face. The general approach was to:

a) give artists pictures of subjects.

b) have artists describe the pictures by the police descriptor set.

c) give their descriptor description to another artist, who would draw the subject based on just the descriptor set.

d) also draw a "photosketch" directly from the photograph.

At this point there was a "photosketch" and "descriptor" sketch for each subject. (Figure 2-4 shows an example of descriptor and photosketch images.) A group of 30 people

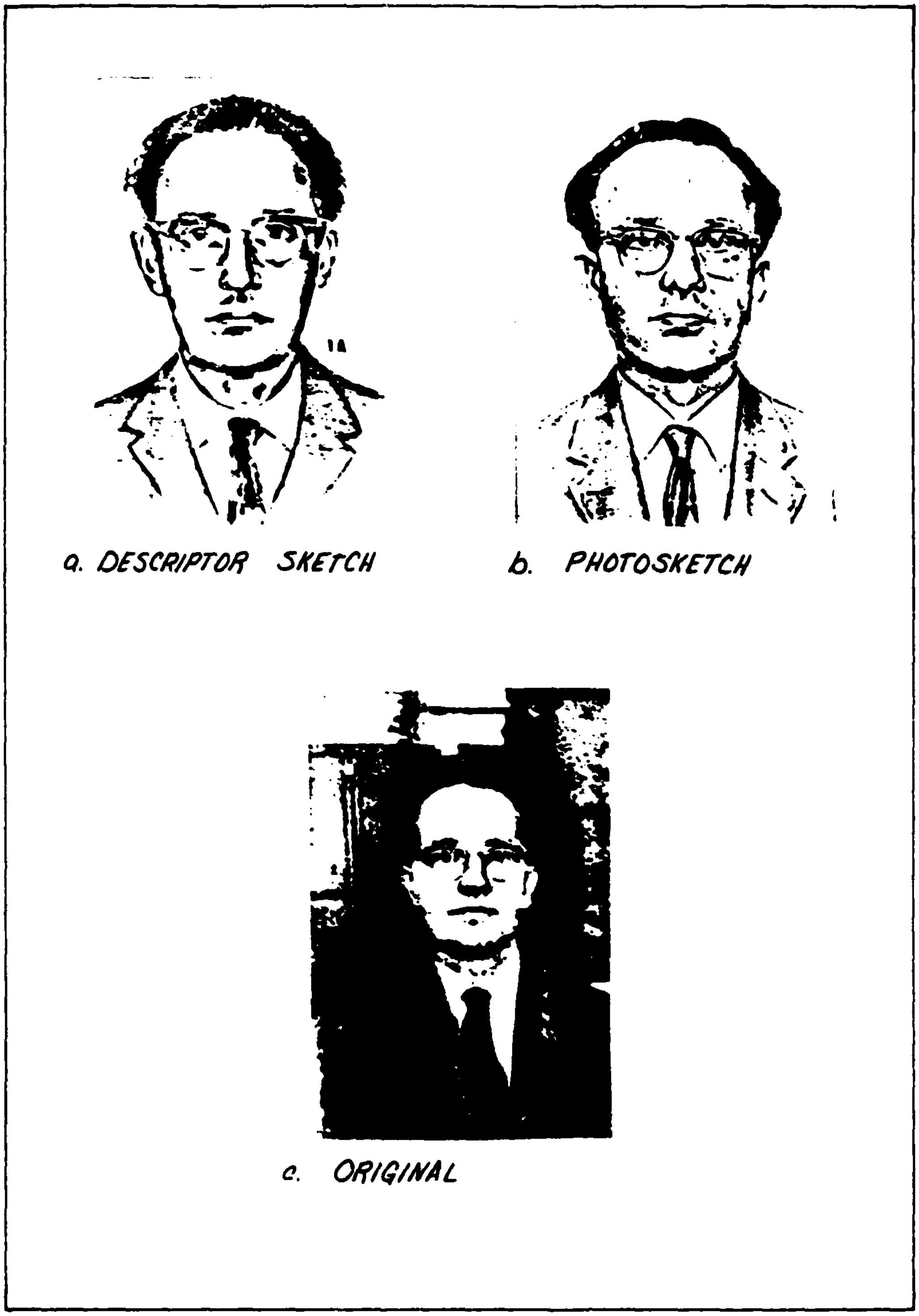


Figure 2-4. Descriptor Sketches and Photosketches
Source: (10)

were tested to see how well they could identify a person out of their group from the descriptor and photosketch pictures. The results: 43% of "descriptor" and 93% of photosketches were identified (10). (I.e., the police descriptor set was not sufficient for good recognition performance.) Other comments included:

a) People preferred lighting of a subject in this order of preference: rear, mixed, and front. Their identification performance tracks their preferences.

b) 80% of the poor recognizers were managerial, while only 33% of the good ones were managerial.

c) Out of 30 subjects, they all displayed differing identification abilities.

d) The subjects doing the identification were also the ones whose pictures were being used. All but one of these identified his own descriptor picture. That person commented that there was something familiar but he could not place the person.

e) Some people are more easily described (and hence recognizable) than others.

Among features thought significant were (in arbitrary order):

a) hair

b) eyes

c) mouth

d) expression

e) suit

f) tie

g) glasses

Harmon conducted a second experiment to see how little information is necessary to represent, pictorially, a recognizable face (10). He reduced 35mm transparencies of faces to pictures with 16 X 16 array elements, with each element quantized to either 8 or 16 gray levels. (An example of such a picture can be seen in Harmon's now-famous picture of Lincoln, shown in figure 2-5.) Preliminary experiments had indicated that a spatial resolution of 16 X 16 was very close to a tolerable coarseness, yielding about a 50% recognition accuracy. Early exploration had also indicated that 8 - 16 gray levels provided recognizable pictures for 16 X 16 pictures. The results were as follows:

a) Overall recognition accuracy was 48%, just under the 50% level sought.

b) Low-pass spatial filtering of the image improved recognition (although this fact was apparently not pursued in the investigation.) (See figure 2-6.)

c) Subject accuracy varied from 21% to 93%.

d) Two out of seven subjects recognized their own pictures "instantly", before the stopwatch could be started. However, they did not recognize the pictures of people

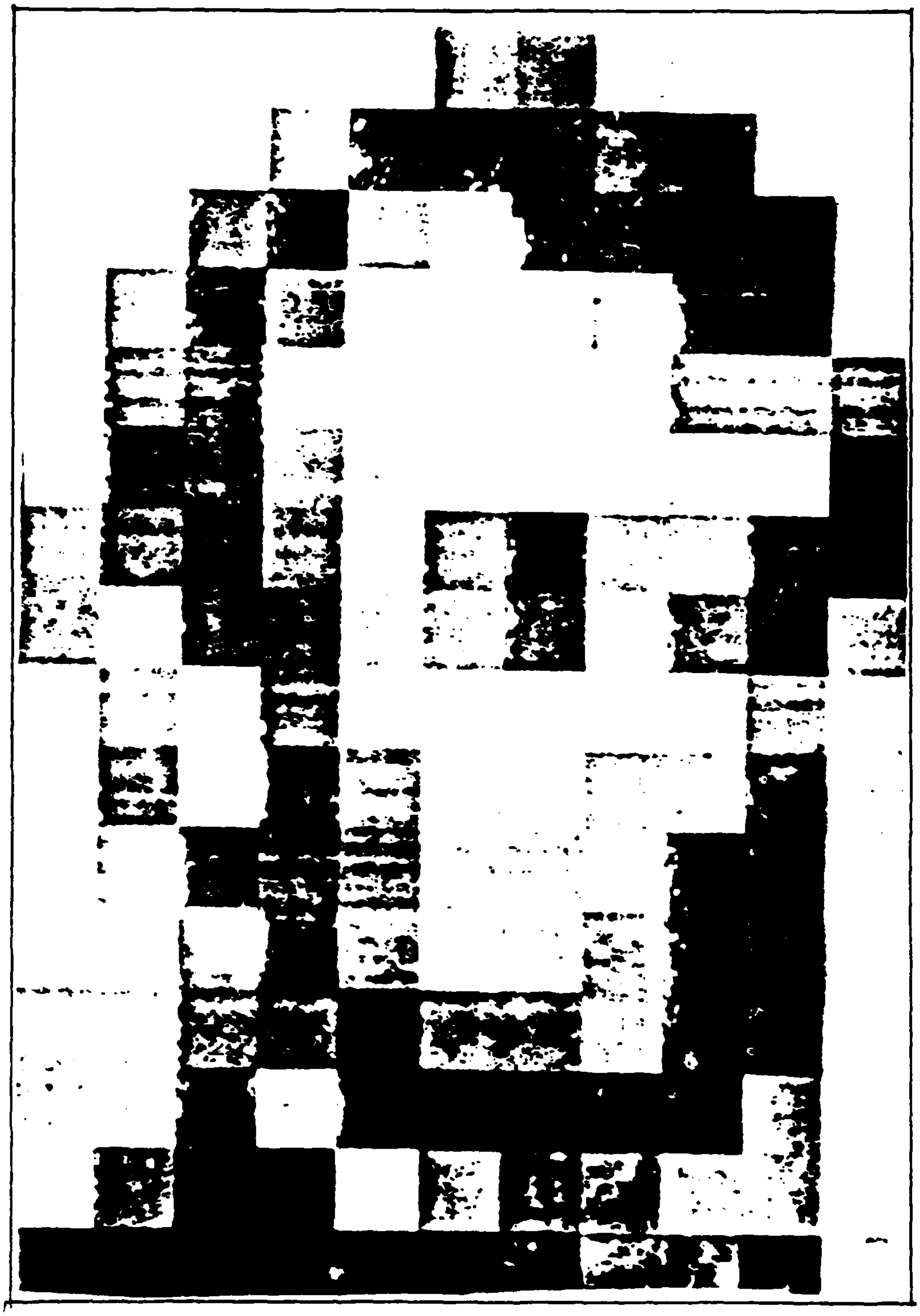
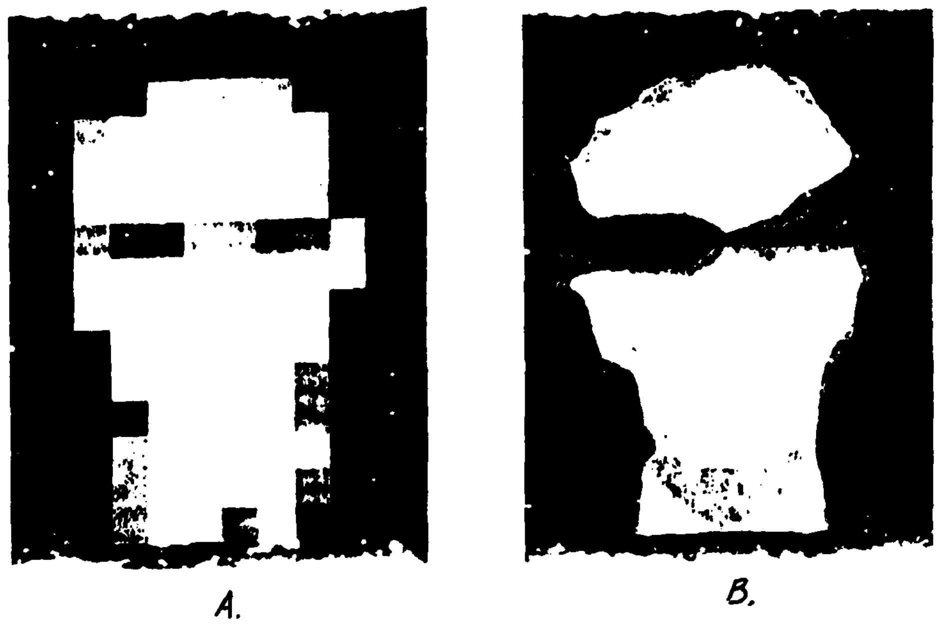


Figure 2-5. Digitized Picture of Lincoln, Source (10)



A. Computer-blurred portrait. Spatial quantization is 16 × 16; brightness quantization is 16 levels. Face is perceived and may be identified at sufficiently large viewing distance.

B. Same portrait as Fig. A, spatially low-pass filtered. Face is perceived and recognized at somewhat less viewing distance than that for Fig. A

Figure 2-6. Effect of Low-Pass Spatial Filtering on Quantized Images
Source (10)

standing next to them.

e) Performance was unchanged with the amount of time available.

Other miscellaneous facts:

a) Motion facilitated recognition

b) Once recognition was achieved, more apparent detail was noticed. Recalled detail affected reconstruction so that once perceived, it was difficult to "unsee" the face.

Goldstein, Harmon and Lesk. In 1971 A. Jay Goldstein, along with Harmon and Ann B. Lesk, published a study in face recognition called "Identification of Human Faces" (7). This group constructed a data bank of 34 facial features (*figure 2-7*) *to be used for identification*, and later reduced them to 22 after eliminating features which were correlated (see Table 2-2.) The study noted at the outset that a human may name and use specific features on demand for a study such as this, but may not in reality be using them in his normal perceptual processes. A theoretical model was developed to relate the number of features needed for identification to the size of the population involved. For instance, (assuming no errors were made!) only two features are required to identify a person out of a group of 50 people, but 10 features are required for a group of 10,000 people. The system, however, did not perform well under actual tests using humans. A person was given a list of

Subject No. 77 — Col. 1-3

Juror No. 6 — Col. 5-6

Experiment No. — Col. 6-8

Date 2/4/69

Item	1	2	3	4	5	Col.
1. [illegible] a. Face Shape	Square	Round	Oval	Long		20
2. [illegible] a. Average	Full	-	Receding	-	Bald	21
b. Length	Short	-	Average	-	Long	22
c. Texture	Straight	-	Wavy	-	Kinky	23
d. Part	Left	Middle	Right	None		24
e. Shade	Dark	Medium	Light	Gray	White	25
3. EYEBROWS a. Weight	Thin	-	Medium	-	Bushy	26
b. Separation	Separated	-	Meeting			27
4. EYES a. Opening	Slit	-	Medium	-	Wide	28
b. Spacing	Close	-	Medium	-	Wide	29
c. Shade	Light	-	Medium	-	Dark	30
5. NOSE a. Length	Short	-	Medium	-	Long	31
b. Nostrils	Narrow	-	Medium	-	Flared	32
c. Tip	Upward	-	Horizontal	-	Downward	33
d. Profile	Concave	-	Straight	-	Hooked	34
e. Projection	Small	-	Medium	-	Large	35
6. MOUTH a. Lip Thickness UPPER	Thin	-	Medium	-	Thick	36
LOWER	Thin	-	Medium	-	Thick	37
b. Lip Overlap	Upper	Neither	Lower			38
c. Width	Small	-	Medium	-	Large	39
d. Nose-to-Mouth Distance	Short	-	Medium	-	Long	40
e. Mouth-to-Chin Distance	Short	-	Medium	-	Long	41
7. CHIN a. Front	Narrow	-	Medium	-	Broad	42
b. Profile	Receding	-	Straight	-	Jutting	43
8. EARS a. Length	Short	-	Medium	-	Long	44
b. Protrusion	Flat	-	Medium	-	Sticking Out	45
c. Lobes	Attached	-	Medium	-	Not Attached	46
9. CHEEKS	Sunken	-	Average	-	Full	47
10. FOREHEAD	Receding	-	Vertical	-	Bulging	48
11. Particularities	[illegible]	[illegible] Eyes	[illegible] Feet	[illegible]	None	[illegible]

(More than one may be selected)

Figure 2-7. Typical Face-Feature Questionaire as filled out by a Juror Source: (7)

Hair	Eyebrows	Eyes	Nose	Mouth
Coverage	Weight	Opening	Length	Lip Thickness
full	thin	slit	short	Upper
receding	medium	medium	medium	thin
bald	bushy	wide	long	medium
Length	Separation	Spacing	Tip	thick
short	separated	close	upward	Lower
average	meeting	medium	horizontal	thin
long		wide	downward	medium
Texture		Shade	Profile	thick
straight		light	concave	Lip Overlap
wavy		medium	straight	upper
kinky		dark	hooked	neither
Shade				lower
dark				Width
medium				small
light				medium
gray				large
white				

Chin	Ears	Cheeks	Forehead
Profile	Length	Sunken	Receding
receding	short	Average	Vertical
straight	medium	Full	Bulging
jutting	long		
	Protrusion		
	flat		
	medium		
	sticking out		
	Lobes		
	attached		
	medium		
	not attached		

Table 2-2. Final Descriptive Features used for Face-Identification Experiments.
Source: (7)

features for an unidentified subject, and from a collection of pictures, was asked to select the picture of the particular subject to which these features belonged. With a set of 128 pictures to choose from, the candidate pictures were reduced to less than 1% for 45% of the trials, and less than 10% for 81% of the trials. The overall study seems to present a promising method for use whenever humans must "classify" pictures, and provides more information on the difficulty and limits of designing face recognition systems. The main drawbacks of this work, however, are that most of the features are very difficult to measure using a computer, and the particular retrieval mechanism used in the study was intolerant of errors.

Psychological Clues to Face Recognition. There have been several studies on the human ability to recognize faces. For instance, a study was performed on the ability of school children of different ages to recognize a whole image from parts of a face (see figure 2-8.) One conclusion was that older aged children more readily identify whole faces from parts. Another was that "various parts of faces differ in how much they contribute to recognition. In addition, parts of faces which have high (or low) recognizability for young children also have high (or low) values for older children (8)." Figure 2-8 also suggests that upper portions of the face are more helpful to

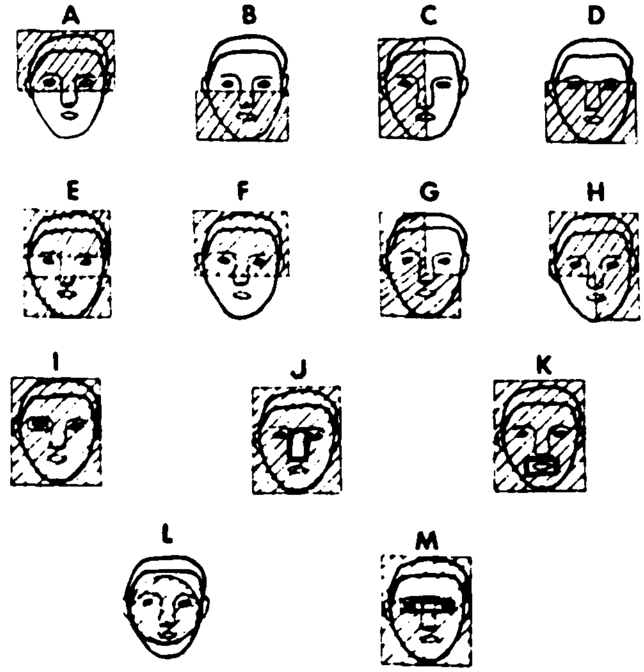


Fig. 1. Schematic representation of experimental conditions. An opaque mask was used to occlude crosshatched sections. Symmetrical features, such as Cond. H, were randomly varied to sample left and right sides.

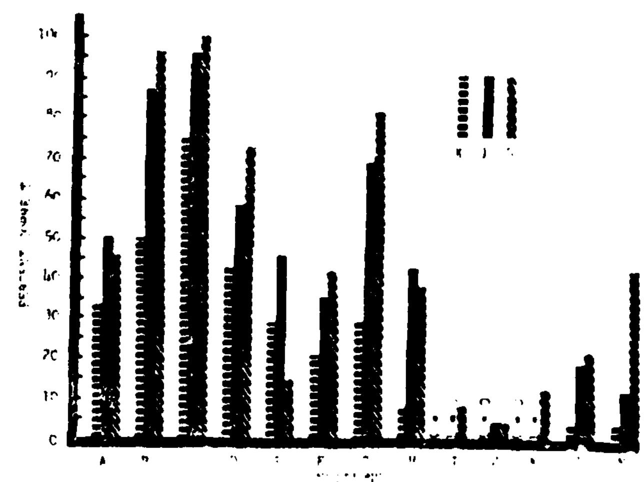


Fig. 2. Per cent correct identifications by kindergarten (K), first (1) and fifth (5) grade subjects on condition A to M.

Figure 2-8. Recognition of Person from Parts of Face
Source: (8)

identification than lower portions (e.g., in the figure, A < B, F < D, and H < G.) Finally, "in a replication of this experiment with length of association as one of the variables, the ability to identify a face on the basis of viewing a part increases with age and is not a function of increased acquaintance."

Another study tested the ability of people to remember previously seen faces (4). This study found the following:

1) Recognition was significantly poorer for inverted faces than faces in a normal position.

2) Recognition accuracy was also impaired when photographs were presented in negative.

3) Recognition accuracy was relatively poorer when comparing two pictures of a person with different expressions ("neutral" and full smile), than when comparing two pictures of a person with the same expression.

A third study tried to characterize what areas of the face a baby paid attention to as it was learning to recognize its mother (9). The experiment measured the frequency with which the baby looked at the mother's eyes, mouth, top of head, and several other areas of the head (see figure 2-9.) The following facts stood out from the study:

1) The eyes were the most frequently looked at (48.9%). The edges of the head was next in percentage (32.7%). The worst performers were the nose (12.7%) and the

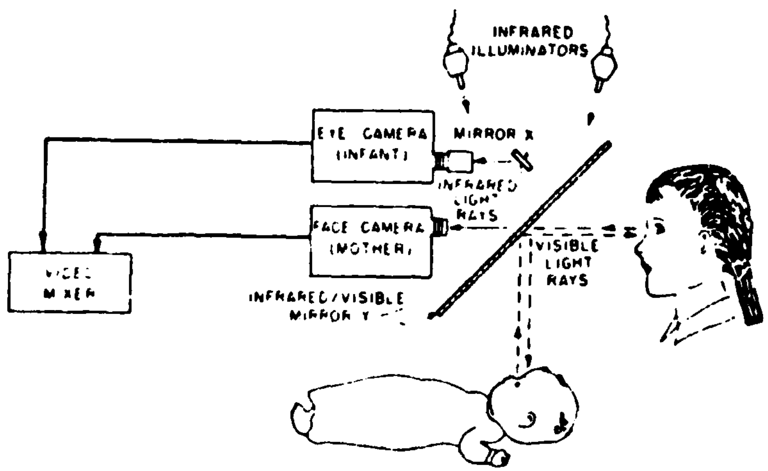


Fig. 1. Schematic drawing of the apparatus used to record scanning of the face by infants. See text and (*8*) for further details.

Figure 2-9a. Eye Contact and Face Scanning in Early Infancy
Source: (9)

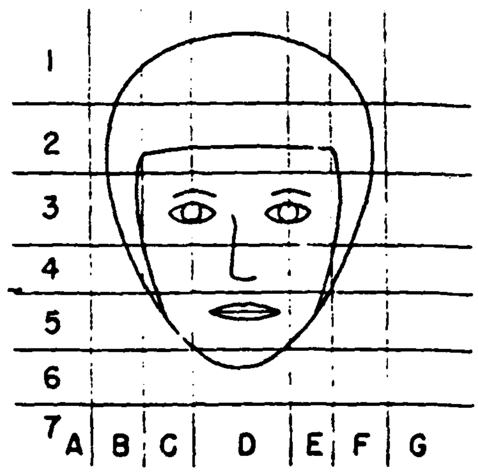


Fig. 2. Scheme of zones for stimulus face. Zones were individually determined for each adult.

Table 2. Effect of conditions on eye and mouth fixations and on the variance of fixations on the face.

Condition	Percentage of face fixations on eyes*	Percentage of face fixations on mouth†	Variance of face fixations along horizontal axis‡ (cm)
Still	50.8	6.0	5.18
Move	49.1	4.2	7.01
Talk	54.1	5.3	4.34

*$P < .10$ by Friedman analysis of variance by ranks test. †$P > .25$ by Friedman test. ‡$P < .01$ by Friedman test.

Figure 2-9b. Eye Contact and Face Scanning in Early Infancy

Source: (9)

mouth (5.7%).

2) When the mother talked, instead of the baby's gaze shifting more to the mouth, it shifted more to the eyes (see figure 2-9.)

If we consider the baby as a pattern recognition machine which adjusts it's attention to the areas of maximum information, this might imply the following:

1) The eyes and the outline of the head provide the baby with the most recognition information.

2) When the mother is talking, the mouth becomes a less reliable source of information, so the gaze shifts to other sources.

Physiological Clues to Face Recognition.

In endevouring to design a vision system consistent with the human system, it would be interesting to study the way an image of a face is displayed on the primary visual cortex. There are two halves to the primary visual cortex, one in the right hemisphere of the brain and one in the left (see figure 2-10.)

The optic nerves from the two eyes converge at the "optic chiasm." The nerves from the left half of both eyes are routed to the right half of the brain, and the ones from the right half of both eyes are sent to the left half of the brain (see figure 2-11.)

Assume we are looking directly ahead at a human face.

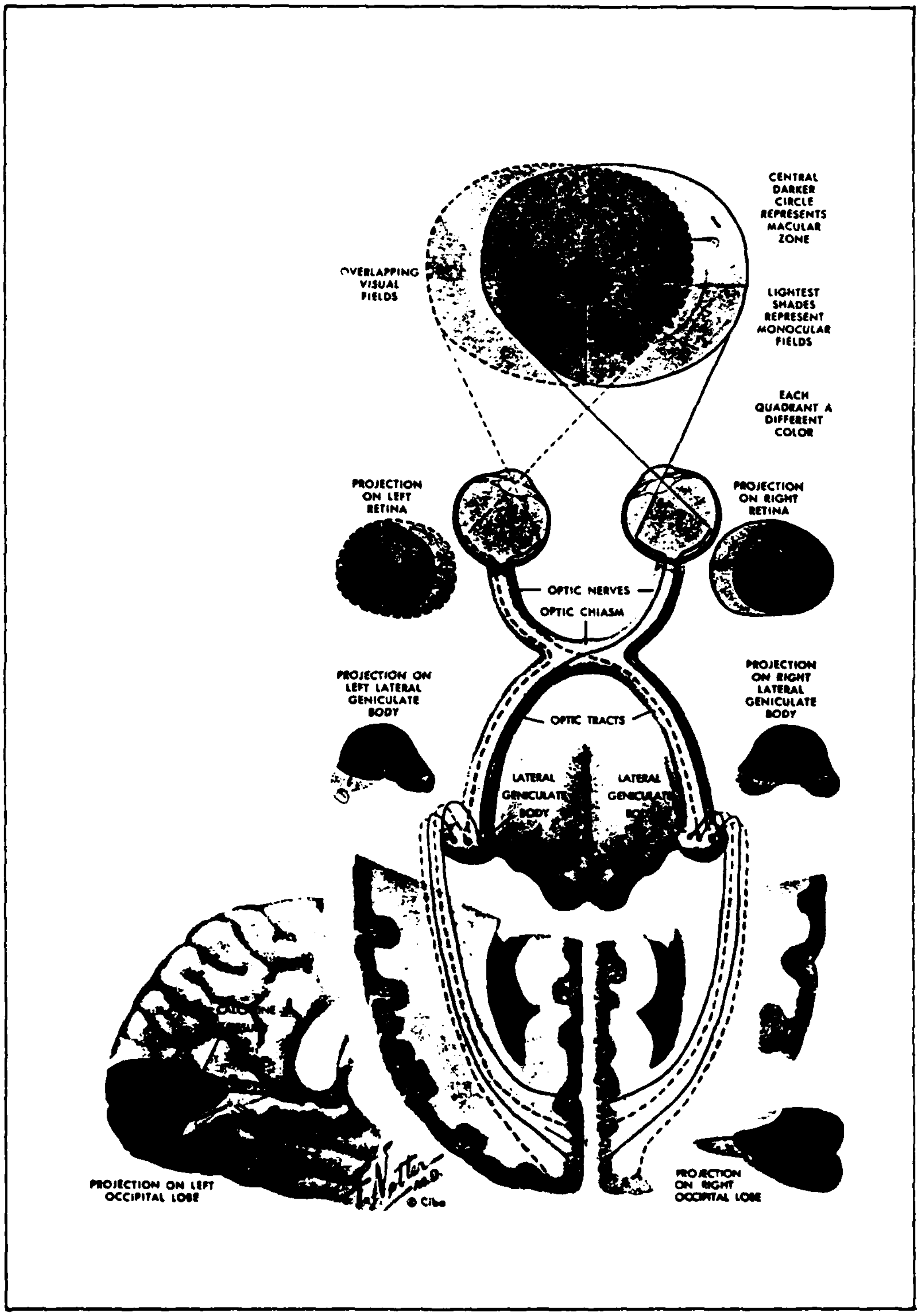


Figure 2-10. Human Eye-Brain System
Source: (18)

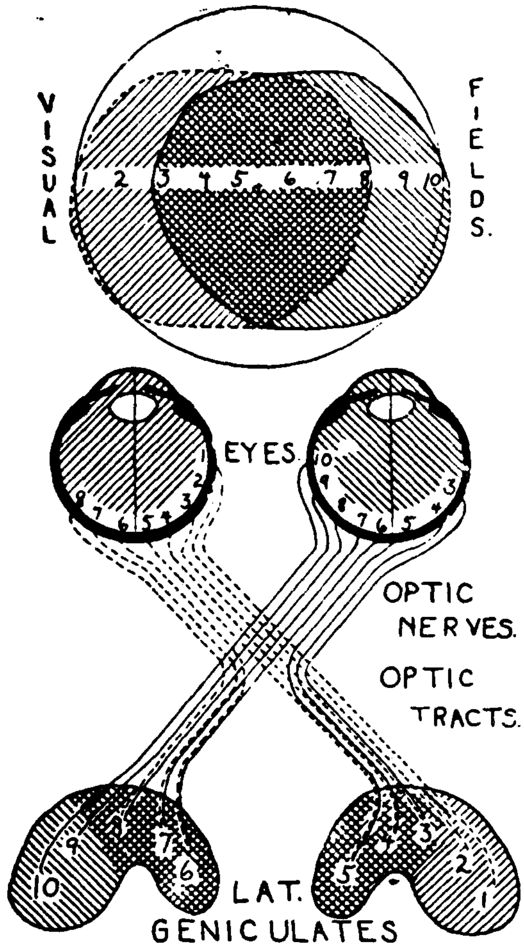


Fig. 112.—SHOWING WHY CROSSING OF OPTIC NERVE FIBERS AT THE CHIASMA IS ONLY PARTIAL.

The major part of the visual field is common to both eyes, but each lateral geniculate body represents just one half of the visual field. Whatever fibers serve any one half of the visual field pass to the appropriate geniculate body. Hence some cross and some do not.

Figure 2-11, Crossover of Optic-Nerve Fibers in Optic Chiasm

Source: (15)

Centering on the eyes, we expect to see two "half-faces", one on each hemisphere of the primary visual cortex (see figure 2-12.) It has not been obvious to researchers why the brain needs to have images divided in this manner. Portrait painters and photographers, however, have long realized that the face is usually asymmetrical, and thus contains different information on both sides.

Any vision system which claims to be consistent with primate vision should be able to process images which have been divided in this manner. If CTT predicted the need for this split-image representation, this would greatly increase our confidence in CTT as a valid theory of human brain function. CTT does, in fact, require such a division, as will be discussed in Chapter 4, "Design of the System."

Features of a "Human-Like" Face Recognizer. The work by Harmon, Bledsoe and others provides a framework for understanding the human abilities for identification of faces. Based on these previous studies, the following features would seem to be associated with a system which identifies faces in the same way as does a human:

a) The system takes multiple looks to investigate different features

1) It needs only a limited feature set (22 at most, according to Harmon)(10)

2) It should only need from 5 - 10 looks to

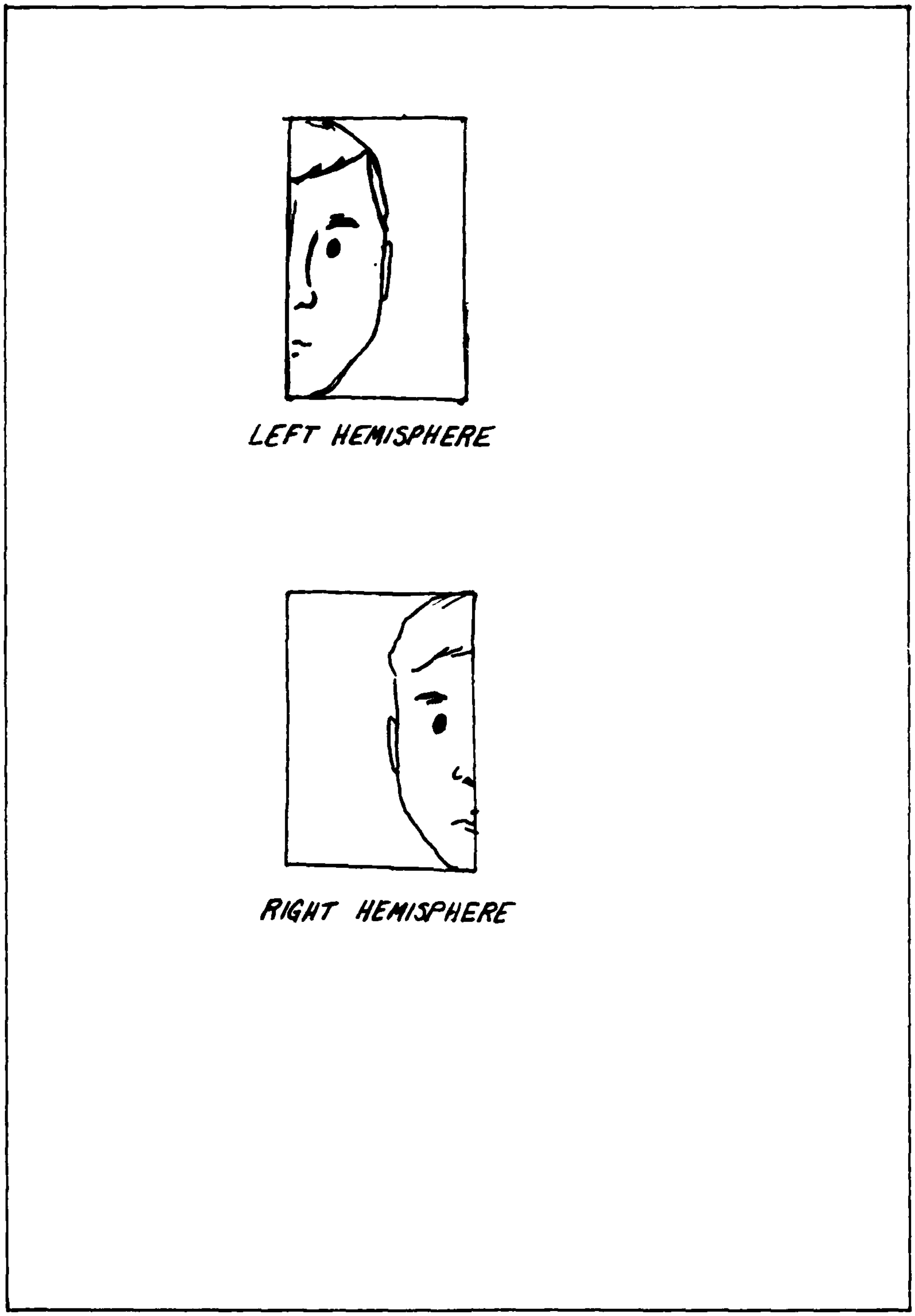


Figure 2-12. Author's Concept of Face Displayed on Left and Right Hemisphere of Brain

identify a person, depending on the number of people in the population (7).

b) As a 16 X 16 array with about 16 gray levels provides about 50% recognition accuracy, then use of about a 64 by 64 array should provide adequate resolution for quality computer recognition. The minimum number of gray levels required is from 8 to 16. (10)

c) The lighting used should be rear or mixed. (10)

d) "Faceness" seems to be inherent in low-pass spatial filtered versions of faces, suggesting the use of Fourier techniques to recognize faces (10).

e) The system should add more detail to an image once the initial identification is made (10). This suggests a mechanism which tries to find the closest match to a subject, but once it does, it "completes the set" with the rest of the stored information about the person.

f) The system should recognize images it has seen often with a reduced number of looks and under increased "noise" or clutter. As mentioned in Harmon's study, most subjects with 16x16 pictures recognized their own faces readily even under reduced data conditions (10). However, as these subjects could not do the same thing for pictures of the people standing next to them, this might suggest that the visual image is first compared against well-trained images for matches, and then subjected to another recognition

scheme afterward for images less well-trained. This also implies that it is not necessarily the recency of the image that is important (because the person whose picture was being examined was sometimes standing right next to the examiner, and yet the examiner did not recognize whose picture it was.) What is important is the frequency with which it has been seen and recognized (our own images are seen about every day in the mirror.)

g) Performance of the recognizer will vary for different people, because some people are more "recognizable" than others (10).

h) The system should experience difficulty with negative images (4).

Having discussed the problems involved in computerized face recognition and the human recognition characteristics found for faces, a framework is provided for better understanding of the chapters which follow.

III. Background of Cortical Thought Theory

Since the purpose of this research is to apply Cortical Thought Theory to the domain of vision, it would be instructive to review the major concepts of this theory. Cortical Thought Theory was the PhD theory developed by Dr. Richard L. Routh while he was at the Air Force Institute of Technology from 1983 to 1985 (21). For years, those involved in Artificial Intelligence (AI) have tried to model human thinking by using logic and other deductive processes. They have enjoyed considerable success in many areas, but computer systems still have great difficulty reproducing what we call "insight", or taking two pieces of information and inducing a new association. Another problem with conventional AI systems is that the search time increases exponentially with the size of the knowledge base.

Rather than starting with basic operations (primitives) using deduction (a concept well-established in AI), Routh approached the problem by starting with primitives of induction. His theory proposes that information is displayed as a two-dimensional image on the cortex. Then the cortex must extract a two-dimensional vector from the image, which he referred to as the "gestalt" of the image. He maintained that the dimension of the gestalt feature vector set must be "two". This type of representation allows direct memory access, which means basically no increase in search time, even with any increase in size of

the knowledge base. This 2-D vector is all that is passed up to the next level of abstraction.

Routh's work also embraces the work of Dr. Leslie Goldschlager from the University of Sydney in Australia. Goldschlager, studying brain theory on an independent course from Routh, explains how a local cortex surface performs the operations of set completion and sequence completion (6). Set completion is an operation in which all points of a set are retrieved, given a unique subset. This characteristic may explain such phenomena as recalling many things about a person seemingly simultaneously, given only the person's name. Sequence completion embodies the AI concept of scripts, in which points are stored in the order in which they occur. Given a unique subset of these points in the right order, sequence completion will retrieve the rest of the points in the sequence.

Combining the retrieval characteristics of set completion and sequence completion with Routh's model, Routh proposed a model of a complete human reasoning system (21).

A joint paper was written (22) explaining some of the salient points of CTT. The contents of the paper are presented in Appendix A, as the level of detail in the paper is appropriate for proper background in this subject. (The details on image recognition are omitted, as they are contained elsewhere in this thesis.)

IV. Design of the System

Part 1 -- DEVELOPMENT OF INITIAL FACE RECOGNITION MODEL

In amateur astronomy, there is a saying which goes something like the following: "When learning to build a six inch mirror, it is easier to build a four inch mirror and then a six inch one than to start out building a six inch mirror." Rather than trying to initially implement the entire CTT vision model, it was first applied to finding only how it mapped faces in CTT space. This section discusses the general CTT model, the initial face recognition model, and finally the analysis of the model and new requirements for improving the model.

CORTICAL THOUGHT THEORY MODEL

The first step in designing a vision machine based on CTT is to examine the general requirements which CTT outlines for a human-like information processing system (see figure 4-1):

1) Display the information as a 2-dimensional image

2) Define the proper boundaries, or "windows", on the image

3) Extract different sub-looks, or "sub-windows", from the image

4) Calculate the gestalt of the different sub-looks

5) Display the gestalts from all the windows as points in a new image

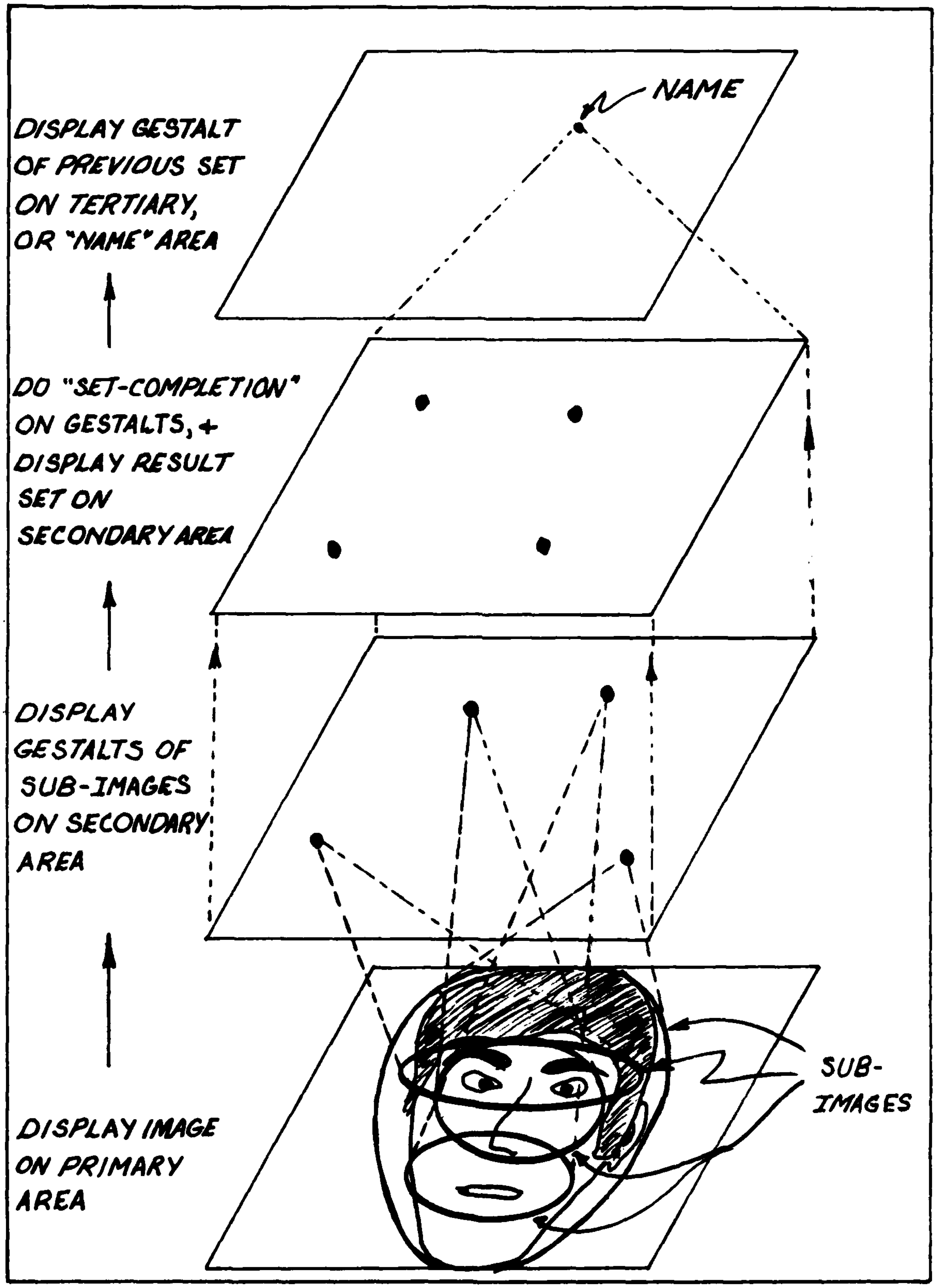


Figure 4-1. Initial CTT Image Processing Concept

6) Apply "set completion" to find the set of previously-seen points to which this new set maps to most closely.

7) Find the gestalt of this new image. This will be displayed as a single point on a 3rd level of abstraction, and is the "name" of the original image.

This results in a surface displaying the names of all the images the system has seen. To "recognize" an image, the system would calculate its 2-dimensional gestalt coordinates. The name of the image is then whichever previously stored point on the "name surface" to which the coordinates of the unidentified image are closest.

INITIAL FACE RECOGNITION MODEL. As a start in evaluating the CTT model for vision, a 64 by 64 pixel by 16 gray level image of a human face was substituted for the image of the audio signal in Routh's speech system, and the gestalt of these images was used to "classify" the faces (see figure 4-2.) The binary values of the image were adjusted so that "white" had the lowest value and "black" had the highest value. The criteria for defining the proper window on the image was to center the face horizontally in a 64x64 pixel box cursor on the screen, and the adjust the zoom on the camera until the top of the head and the bottom of the chin just fit within the top and bottom of the box cursor (see figure 4-3.) White cardboard was used as a background for the pictures (see figure 4-4.) The pictures

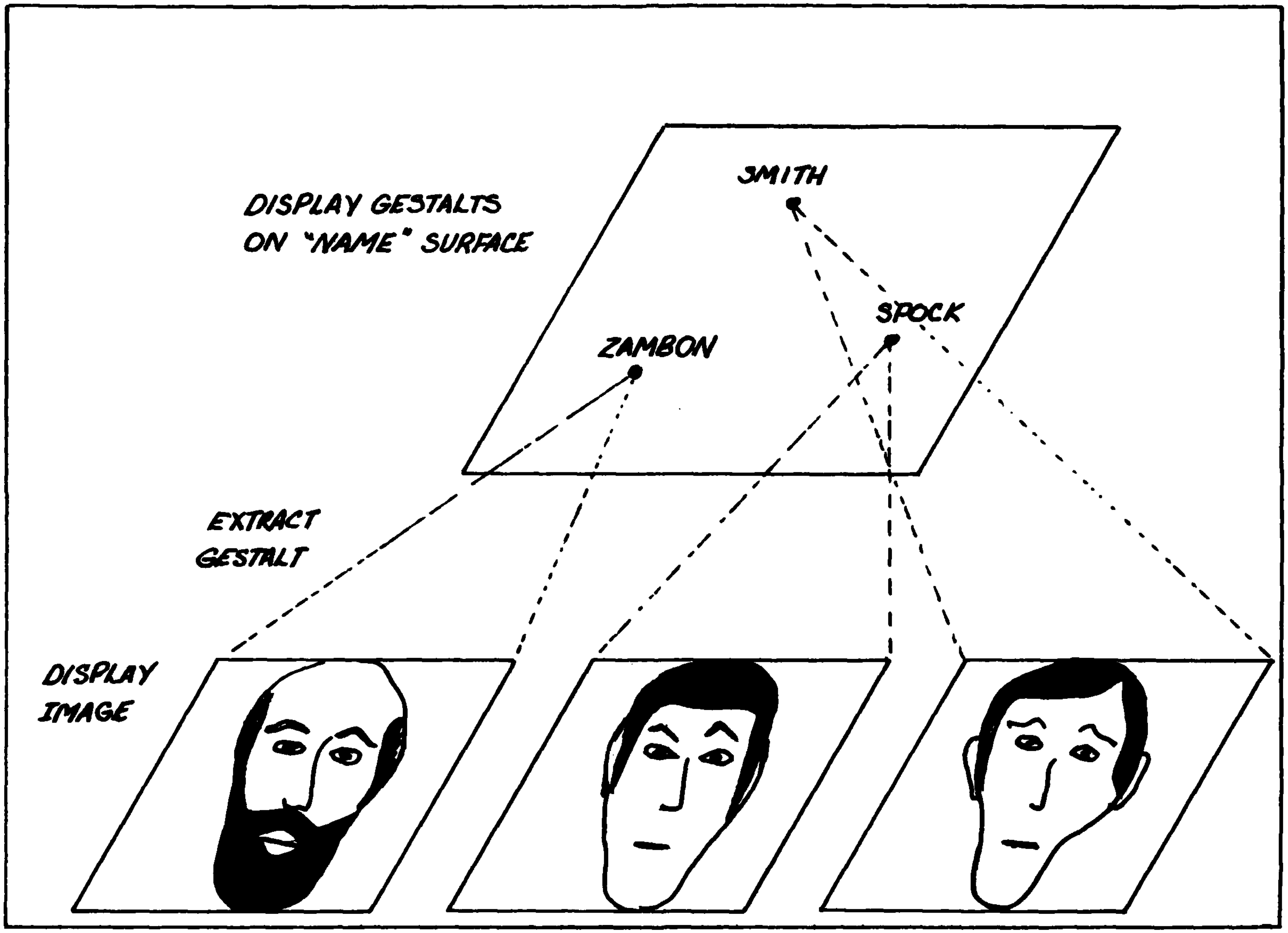


Figure 4-2. Initial CTT Face Recognition System

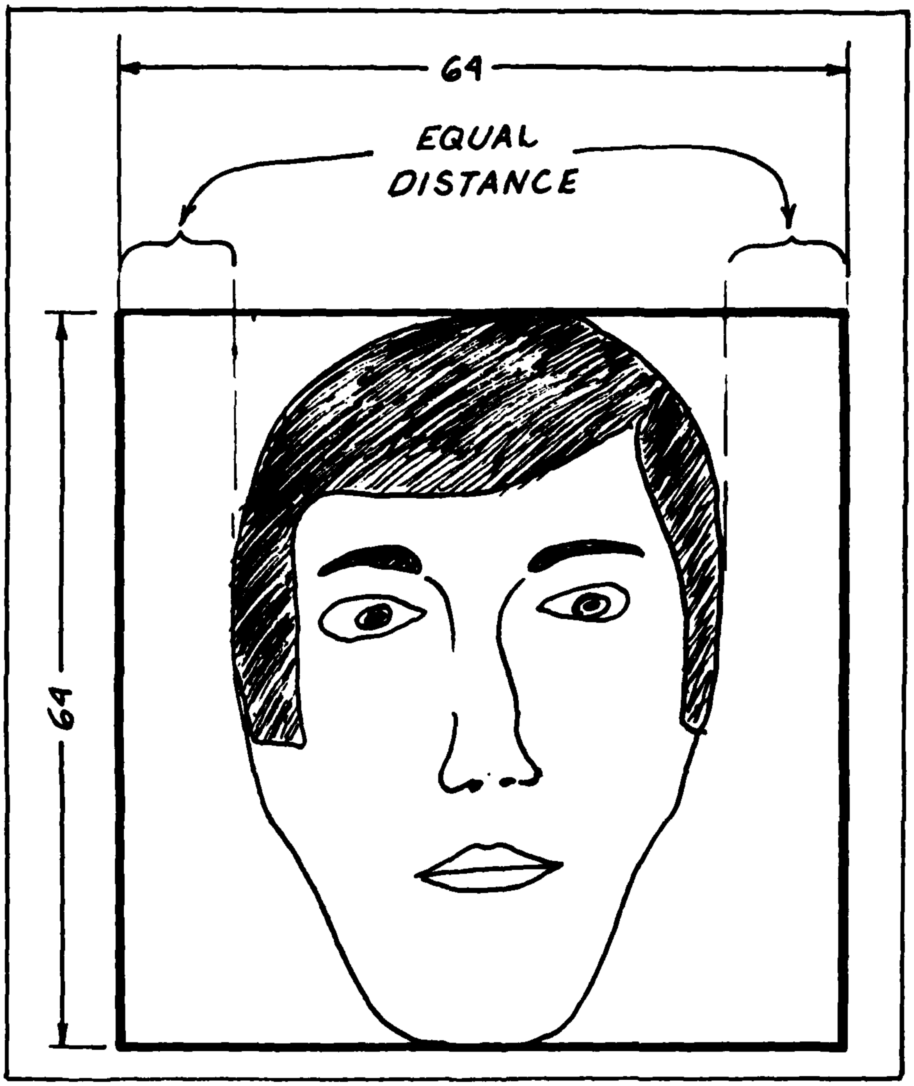


Figure 4-3. Windowing of Face within 64x64 pixel
Box Cursor

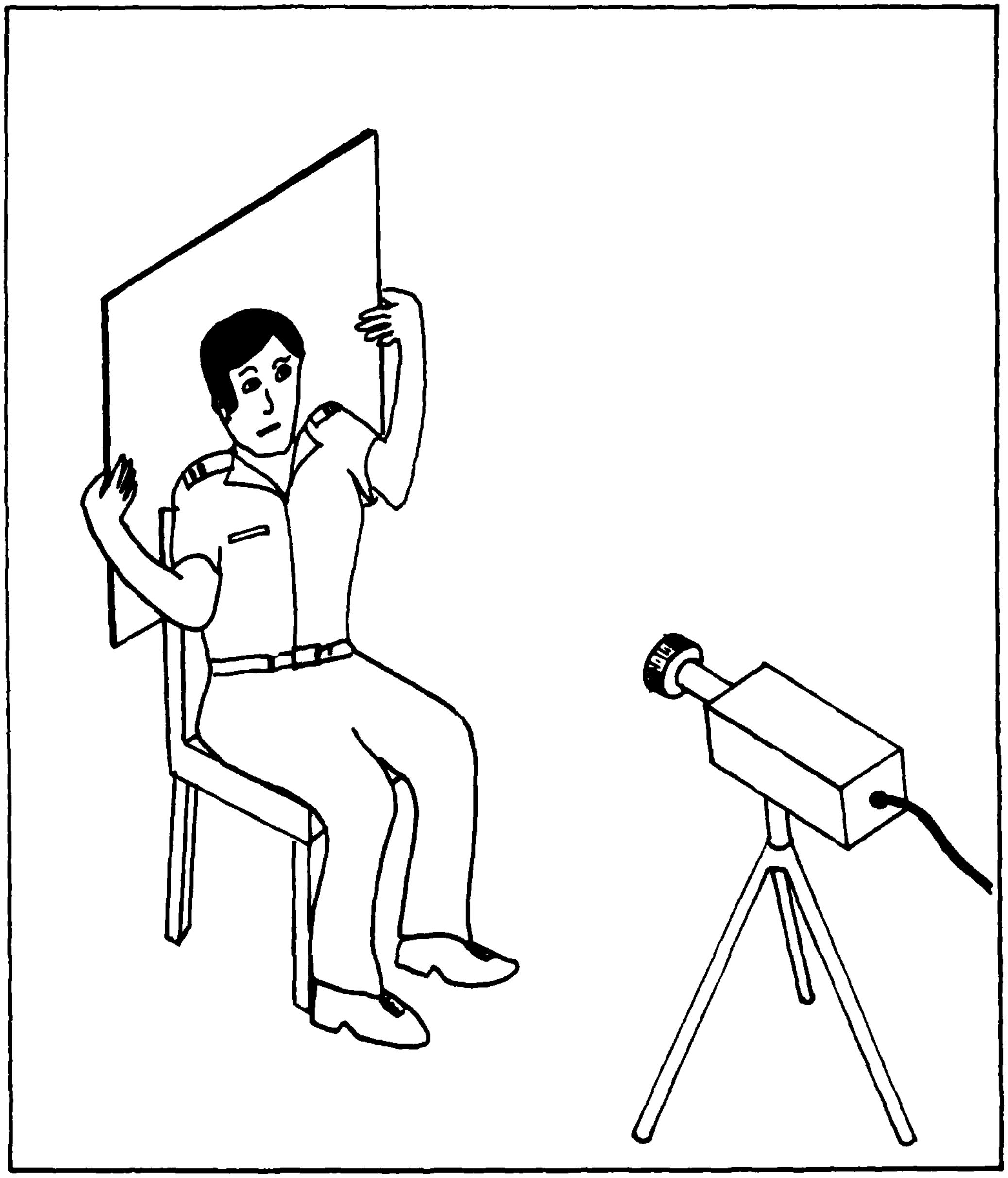


Figure 4-4. Setup for Initial Testing

were taken so as not to show the shoulders. The same gestalt mechanism used to process the audio signal was used to find the gestalt of the human face.

ANALYSIS OF INITIAL RESULTS. Significantly, different faces could be distinguished by this method, as shown in figure 4-5. These results indicated the following:

1) Human faces can be classified and distinguished with the Routh CTT model.

2) 10-15 faces can be reasonably identified using one plot as in figure 4-5. However, the plot quickly becomes crowded.

3) People with beards and/or mustaches clustered to the left side of the plot, while people with a lot of dark hair on top of the head and no lower facial hair clustered to the right.

In addition, these results revealed several new requirements for an advanced face recognition system. These will be discussed under the following categories:

1) Calculation of a Gestalt,

2) Windowing Mechanism, and

3) Contrast expansion.

PART 2 -- DEVELOPMENT OF AN ADVANCED MODEL

The advanced face recognition model is discussed in three sections:

1) Facial image processing -- processing on a human face image necessary to characterize a face as a set of

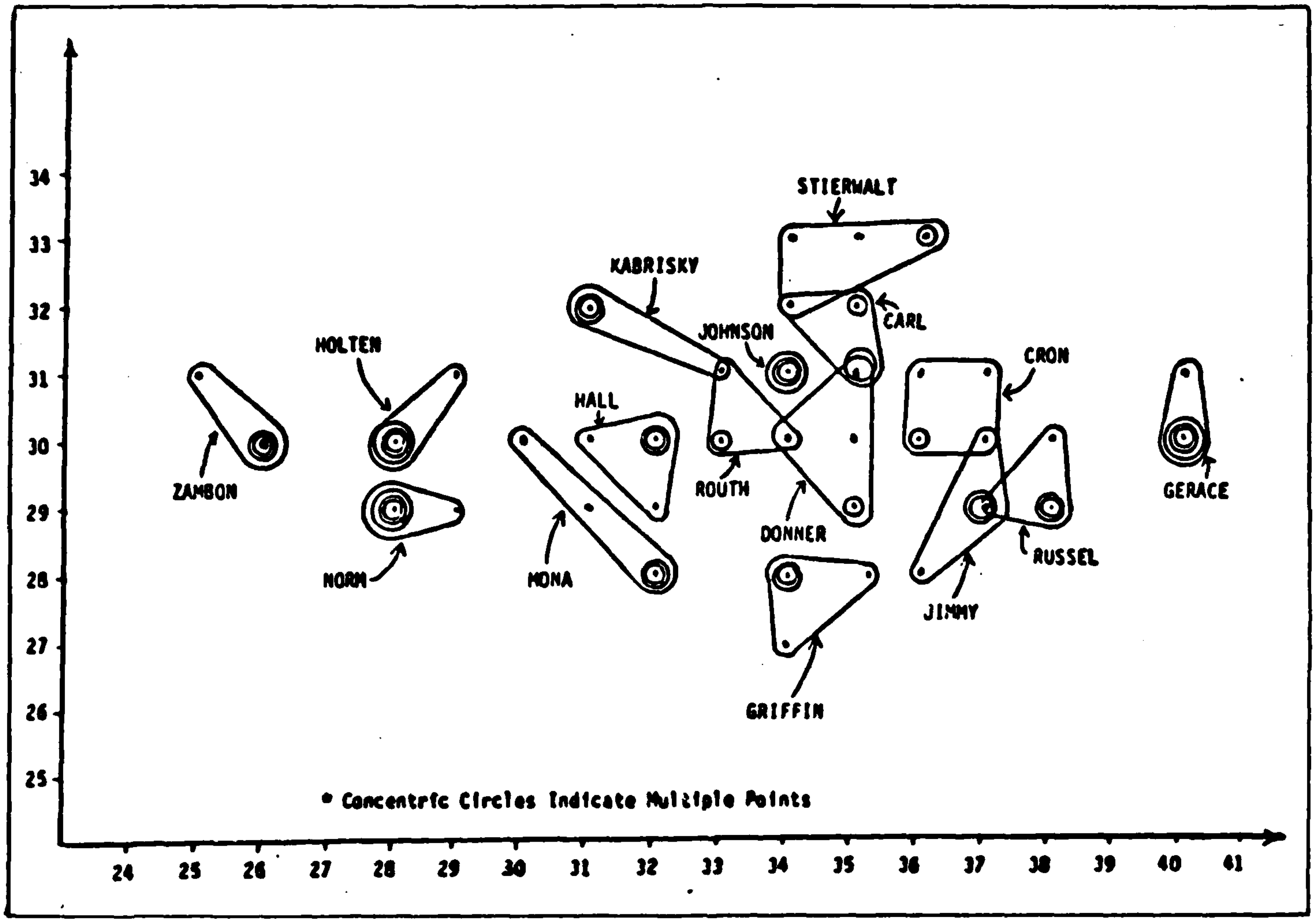


Figure 4-5. Full Face Gestalt Plots (Initial Results)

gestalt coordinate points.

2) Learning a face -- processes involved in training system with several images of a person, extracting statistics from the data, and generating a database which is the "facial memory."

3) Identifying a person -- how the search space in memory is directly computed using CTT, and how the closest match is found using a distance metric based on set completion.

FACIAL IMAGE PROCESSING

There are three processes involved in processing a human face: calculation of gestalt values, contrast-enhancing the image, and picking proper windows on the face for the gestalt calculations. These are discussed below.

CALCULATION OF A GESTALT

The gestalt transformations used in this research are described in Appendix A, equations 1, 1a, 2, and 2a. (Those interested in more detail concerning these transformations should consult Dr. Routh's dissertation (21).) Several different issues concerning optimizing the use of the gestalt transform are discussed below.

a) Processing for Scale Invariance. Humans can recognize an object irregardless of scale. This feature was incorporated as part of the basic processing of the gestalt calculation, and is one of the expansions made to CTT for the visual system. It is accomplished by calculating the

gestalt of the original image, and then expanding the value of the X & Y coordinate values of the gestalt point until it is where it would have been if the image had been full size.

The scaling is done as follows:

If (X',Y') = Original Gestalt value

then the New Gestalt = (X,Y), where

$$X = X' * 64/A \tag{4-1}$$

$$Y = Y' * 64/A \tag{4-2}$$

and $$A = \max(Wx, Wy) \tag{4-3}$$

where X,Y Window values are the size of the original image.

This process is illustrated in figure 4-6.

b) Correction of Blind Spot. Another characteristic discovered of the original gestalt transform was that it was basically blind on the left 16 columns and the top 16 rows (see figure 4-7.) To understand this, it is necessary to look at the way the gestalt transform is calculated (see figure 4-8.) The original gestalt transformation used by Rought did a point by point correlation of each row of the input image with the 64 spatial sub-harmonics from 1/64 to 1, by increments of 1/64 (Appendix A and reference 21.) The resulting correlation

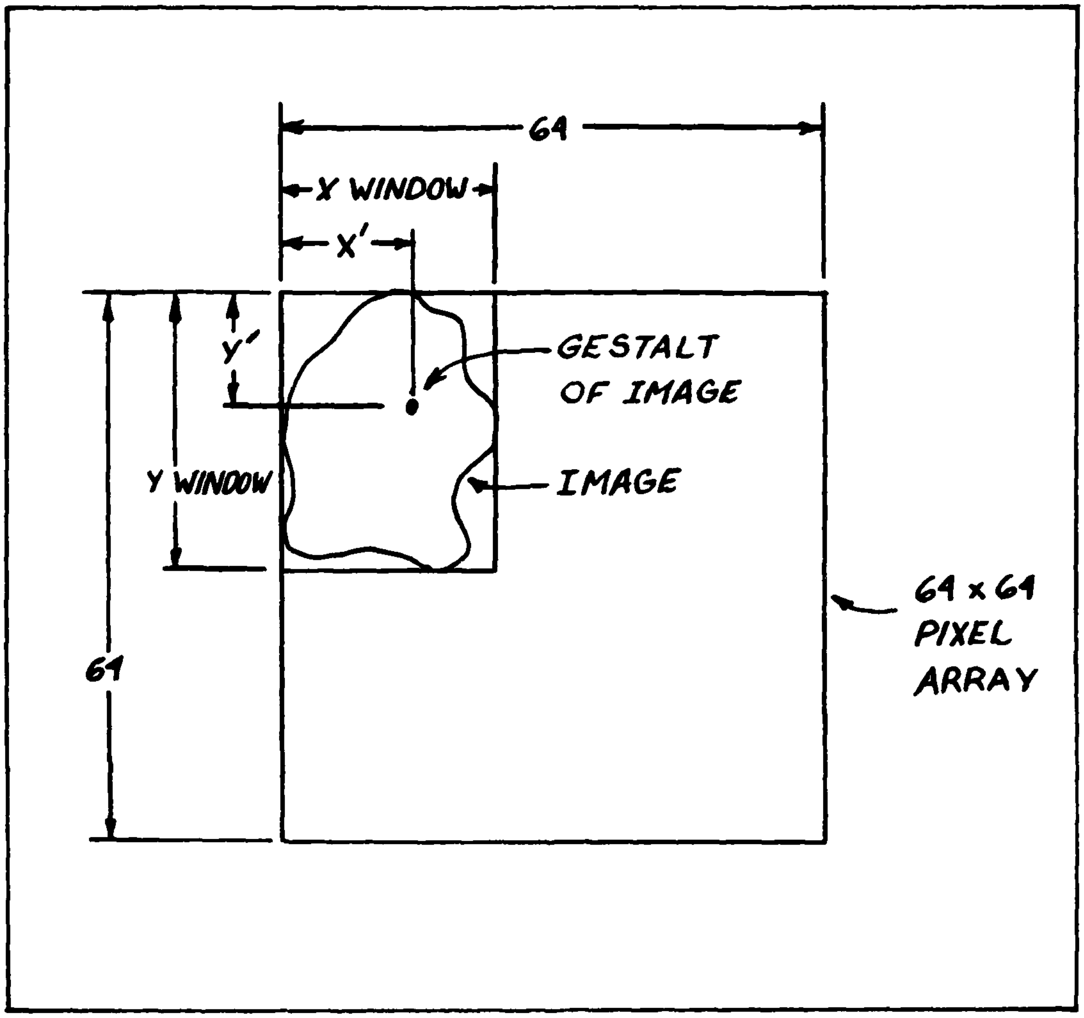


Figure 4-6. Dimensions Used for Scale Invariance Calculation

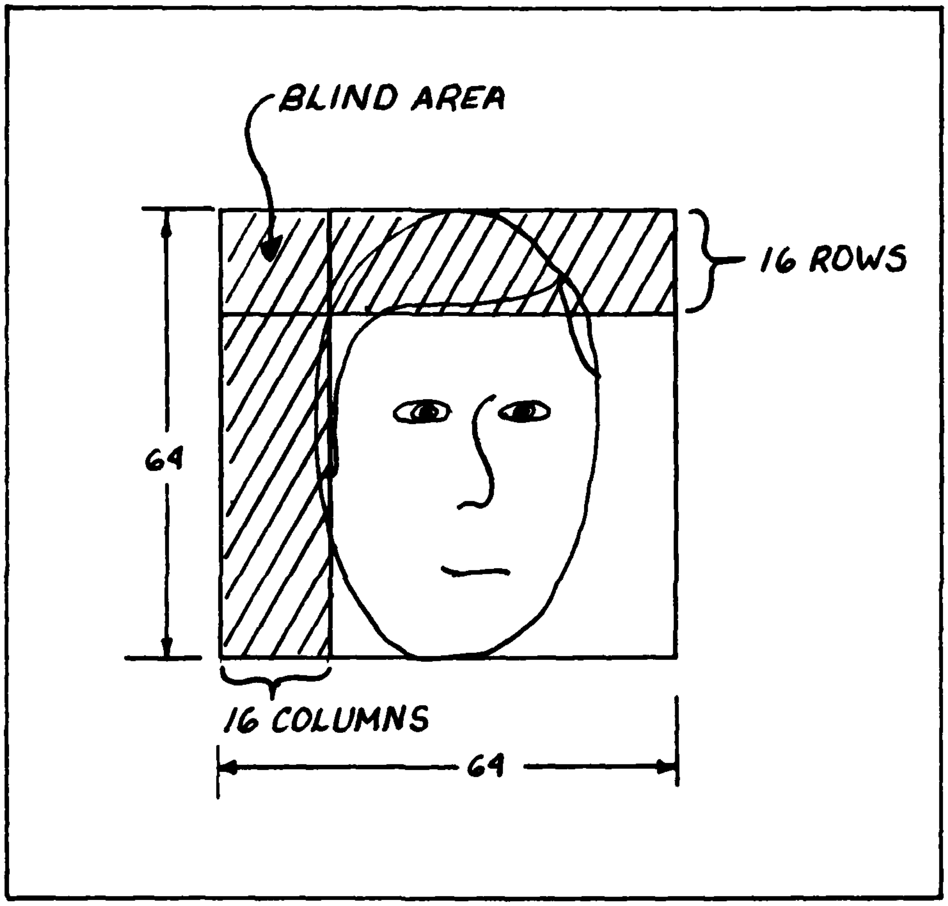


Figure 4-7. Blind Area of Original Gestalt Transform

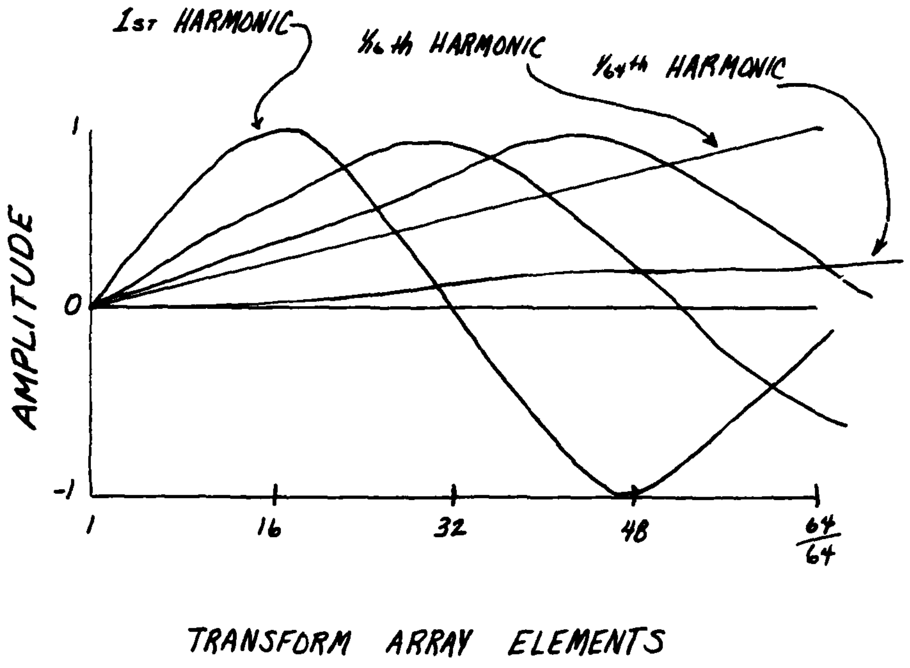


Figure 4-8. Spatial Sine Waves with which an Input Waveform is Correlated in Routh's Initial Gestalt Transform

values were then substituted back into the image array row, the element 1 getting the value from sub-harmonic 1/64, the second element getting the value from sub-harmonic 2/64, and so on, until element 64 received the value from the first integral harmonic. This process was continued for each row. The process was then performed on the columns of the resulting 2-dimensional array, in the smae way as it had been done for the rows. The problem resulted from the fact that none of the peaks from the sub-harmonics extended below the 16th array element, as seen in figure 4-8, meaning that the gestalt coordinate values less that 16. This meant that the hair and the left side of the face had diminished effect in the gestalt calculations. The question might be asked, "Why not use higher-order harmonics, thus shifting the peak lower on the array?" Routh showed that in order to have the transformation maintain the effect of a "gestalt", (i.e., only ending up with one "hump" in the output image) the transform cannot use greater than the first harmonic of a spatial sine wave (21). For these reasons, a new transform was designed which filled in the blind area, provided a reasonable approximation of the previous gestalt transform, and could reasonably be implemented by the structure of the cortex. Expecting better performance now that the hair was being seen, it was surprising to find the faces grouping closer together! It became clear that what was causing the groupings was not similar-looking facial features, but

similar hair shape and mass. It became evident that to get a better separation, the system needed a look at the face without the hair, plus a look at other sub-parts of the face. The whole face gestalt still provided useful, but not sufficient information.

WINDOWING MECHANISM. Given that a series of sub-looks, (or "windows") on the image might be required for increased discrimination, one might ask,

1) "How does CTT handle this apparent need to process different windows on an image?

and

2) "What are appropriate windows to use?"

Routh recognized the need for specifying a series of reproducable windows on an image being processed by his gestalt mechanism. CTT proposes that the eyes might automatically be calculating window locations for several areas of greatest contrast in an image, and supplying these locations to the primary visual cortex. Routh proposed a mechanism by which this calculation might be being performed in the eye by the retina. However, it would apparently be extremely difficult to implement with a conventional computing architecture (21). An approximation to the proposed retinal windowing process, however, was developed for the domain of human faces. In addition, it was determined that the facial images must be split vertically down the center before processing.

PROCESSING SYMMETRIC IMAGES -- THE NEED FOR HALF-FACES. Since the gestalt transform tends to find the center of mass on an image, there are some problems which will be experienced when using this particular transform. One of the worst is that the system is not sensitive to aspect ratio. For instance, look at figure 4-9.

If a face is symmetrical, then a wide face will give the same gestalt as a thin face. Unfortunately, people tend to be quite aware of aspect ratio when recognizing someone (determined by an informal survey by the author.) The previous plot of whole faces reflects this problem in the small range of X values vs the relatively large range of Y values.) In addition, the system can't tell the difference between a woman with long hair on the sides and a man with thin hair on the sides.

To handle this problem, it was necessary to divide the image down the center, display the halves as two separate images, and take the gestalts of the separate images (see figure 4-10.) Now changes in aspect ratio are reflected as changes in the X direction of the gestalt.

The author, wanting to be consistent with CTT and the physiology, found this split-image requirement to be a strange restriction of the presentation of a facial image. Then he realized that the primate visual system splits images vertically down the center before displaying them on separate left and right primary visual cortexes (as

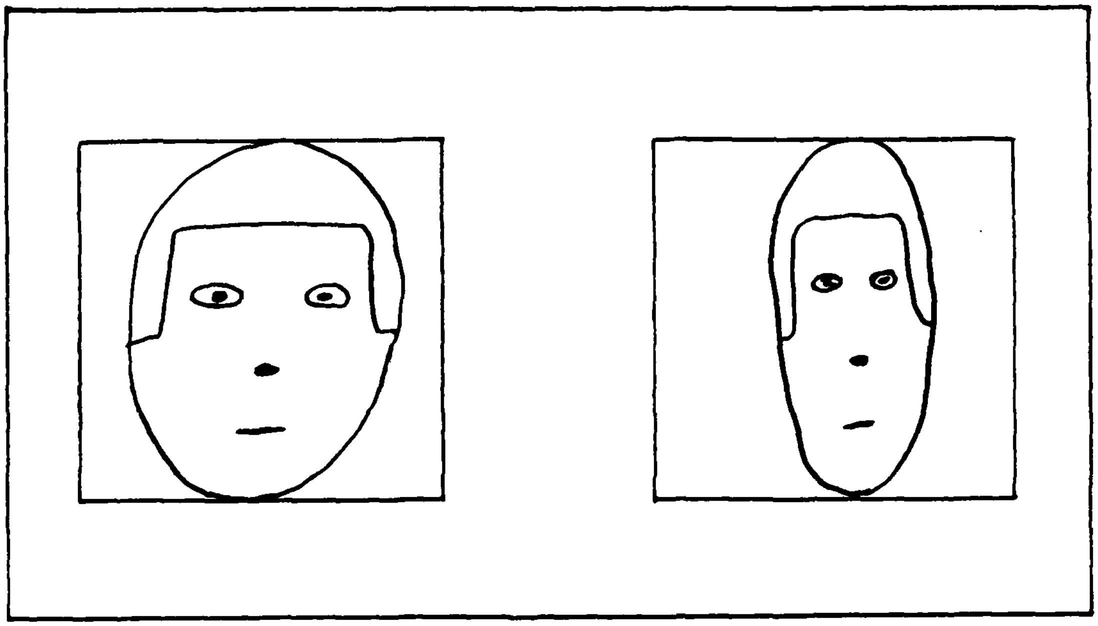


Figure 4-9. Example of Faces with same Gestalt Value

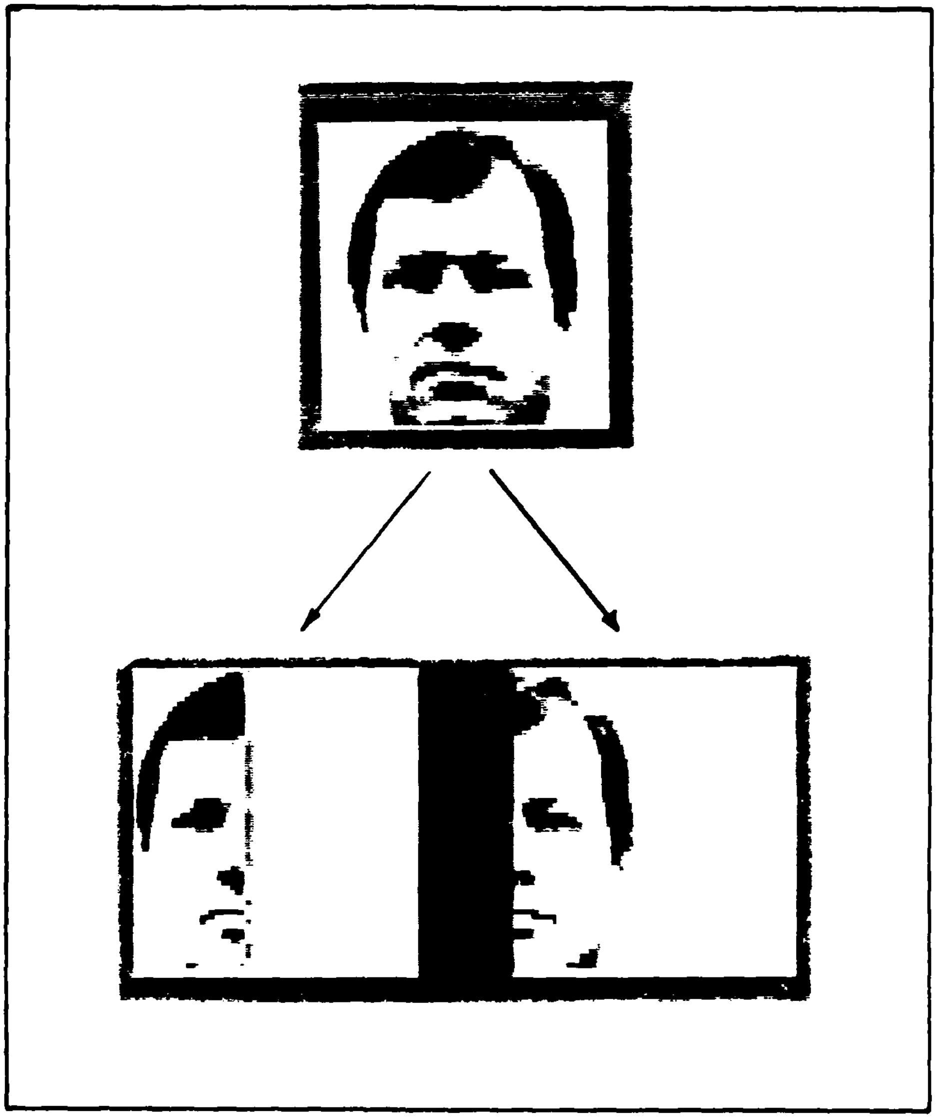


Figure 4-10. Division of Face into Two Half-Images

discussed in chapter 2.) The reasons for the partial splitting, (or "decussation") of the visual pathway at the optic chiasm are not well understood, and attempted explanations for the phenomenon quickly become complex and convoluted. It is significant than Cortical Thought Theory provides a possible explanation which is simple, straightforward, and is a natural requirement of the theory. It shows that the reason for the splitting is that it's needed to provide higher-quality form-discrimination among vertically-symmetric forms.

DETERMINING PROPER SUB-WINDOWS. As previously mentioned, the actual process which CTT predicts that the retina uses to find windows is too complex for present architectures. However, a simplified process was determined for the domain of human faces. The facial image is first contrast expanded to emphasize the high-contrast areas of the face. Then straight lines are used to mark the boundaries of the different significant facial features, resulting in a plot as shown in figure 4-11. Now, with such a plot, calculation of the retinal window regions predicted by CTT is reduced to finding different combinations of boundaries using the lines.

Which windows should be used? If CTT is correct, the brain may be using scores of them. To limit the problem, the author took the six combinations which seemed most obvious to him. In actuality, many different windows need

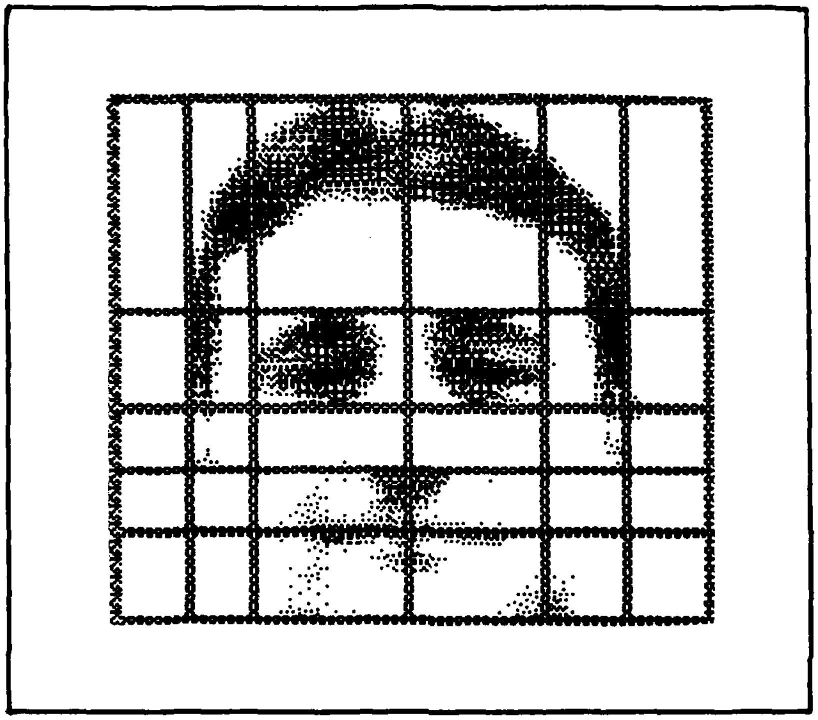


Figure 4-11. Feature Locations used to find Windows on Face

to be tested to determine which ones give the most information about the face. The ones picked initially for this study were:

1) Whole head -- to get separation by hairstyle. Many times, if we are searching for a person at a distance, the first thing we will recognize is the outline of their hair, for the rest of the features may not clearly be visible.

2) Top of eyes to chin -- to look at the face independent of the hair. As a person's hairstyle may change slightly day to day (or for a woman, may change drastically if she puts it up or lets it down), there needs to be at least one window independent of the hair.

3) Top of eyes to bottom of upper lip. When a person moves their mouth, their gestalt can change drastically due to the great potential change in dark mass in the mouth area. To gain some independence of mouth movement, a window was taken from the top of the eyes to the bottom of the upper lip. (When taking a picture in this study, the subject was told to keep their mouth closed, so the bottom of the upper lip was designated as the center of the mouth.)

4) Top of nose to bottom of chin -- used to recognize a person by their mouth.

5) Top of head to bottom of eyes. This window is independent of nose or mouth.

With these windows defined on the face, the system is now able to extract portions of a face in a repeatable

manner, and calculate their gestalts. In addition, this makes the system shift-invariant, as the window boundaries move as necessary to find required features.

CONTRAST ENHANCEMENT

When initially taking pictures and processing gestalts, the effect of lighting and f-stop needed to be evaluated. A Dage video camera was used in this study, and included adjustments for f-stop, focus, and zoom (see chapter 5 for specifics.) The normal lighting in the lab area was used, as it was provided fairly even illumination from rows of overhead lights oriented parallel along a line from the camera to the subject. (The equipment and studio setup are shown in figure 5-2.) For simplicity sake the lighting was assumed consistent. To evaluate the effect of different f-stops and determine a correct setting, pictures were taken at various f-stops and their gestalts found. The results were as follows:

1) F11 and above -- the faces were too dark to process features

2) F8 -- the gestalts gave very poor separation, with most of the resulting output regions for different individuals overlapping.

3) F5.6 -- Reasonably good separation between faces of caucasians -- used for subsequent processing. Too dark, however, for dark-skinned people such as blacks.

4) F4 -- Too light to use for light-skinned people, but

best for dark-skinned people.

F5.6 shots, for caucasians, give a high-contrast image in which facial lines are bleached out (for the most part) and hair, eyes, nose and mouth appear as dark blobs. The person is usually still recognizable in this form (see figure 4-12.)

Immediately the question is raised, "Why do we get better separation with a "poorer" image?" It would seem that the location and size of the features left in this contrast-expanded image contain the essential information of facial recognition. Indeed, skilled artists are able to create a recognizable face with just a few brushstrokes showing the eyes, nose, and mouth.

It does seem profitable, therefore, to contrast-expand the images before processing for gestalts. However, as has been noted, F5.6 doesn't work on everybody (e.g., a dark-skinned person.) In addition, the locations of the boundaries of a person's head are required for further processing, but they may disappear when using F5.6.

The answer settled upon was to take the pictures at F8 (where all head boundaries are still visible to the human operator and computer), let the computer extract boundary information from this picture, and then artificially expand the contrast to the proper value. Quantifying the "proper value" is the problem.

Statistical measurements were taken on pictures that

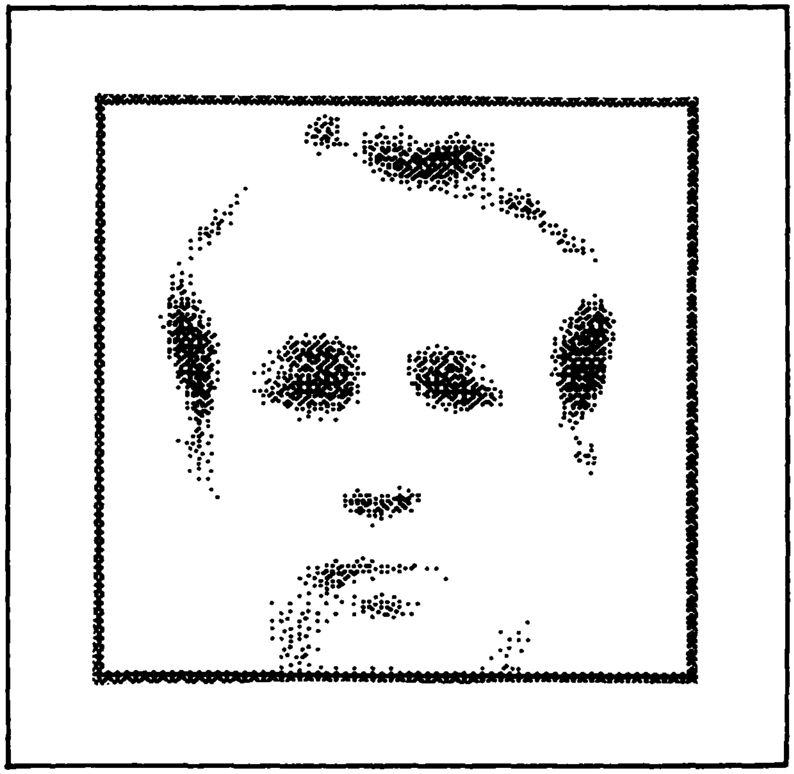


Figure 4-12. Typical F5.6 Picture

"looked" correctly contrast expanded. There was no clear pattern between the mean or standard deviations of pixel values for pictures of the entire face. (This was due primarily to different distributions of hair in different pictures.) It was noted at this point that we seem to be able to see a person's eyes clearly when we are looking at them face-to-face, even when a picture taken by a computer leaves the eyes in shadow. It was postulated that the human visual system might be expanding the contrast around the eye area, since we tend to look someone in the eyes when trying to identify them.

Statistics were taken on just the eye area, but the results were still not consistent across pictures that "looked right" to the author. However, the area between the bottom of the eyes and the top of the nose had a consistent characteristic in all of the pictures -- it was always nearly completely white!

A system was developed which expanded the contrast of the entire picture until the area vertically between the bottom of the eyes and the top of the nose and horizontally between the outside of the two eyes just turned white. This gives quite consistent results, and gives the system a reasonable independence of skin color (see figure 4-13.) The only problem noted so far is that dark-rimmed glasses across a light-colored face may impinge upon the area being sampled, hurting the expansion. Light-colored glasses or

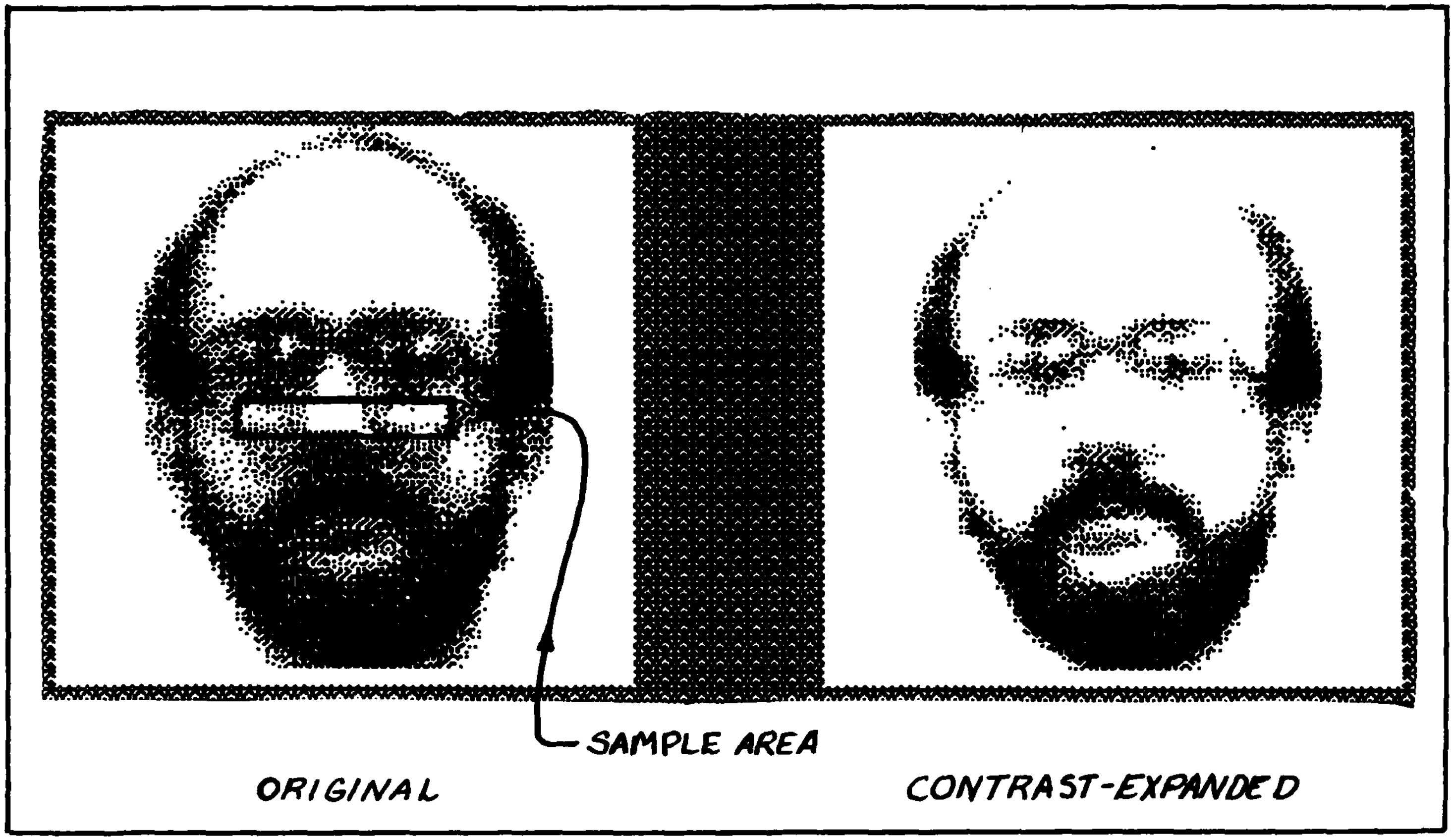


Figure 4-13. Example of Computer Contrast Expansion

wire-rimmed glasses do not significantly hurt the expansion.

Other areas of the face were also candidates, as they too become "white" when properly expanded. These, however, were rejected for the following reasons:

1) Forehead -- not always available to sample due to hair across the forehead.

2) Mouth, cheeks and chin -- not always available due to mustaches & beards.

Although probably not optimal, the contrast-enhancement mechanism described above does a reasonable job in contrast-enhancing facial images for any color skin to the proper value.

SUMMARY OF FACIAL IMAGE PROCESSING. The gestalt transform was given increased resolution and scale invariance. The need for using vertically-split images was discussed, along with how this need provides a possible explanation of the need for partial decussation in the primate visual system. A method was developed for finding the significant windows on the face which CTT's proposed retinal windowing process would have found, and in the process made the system shift-invariant. Finally, a process was developed to properly contrast-expand different facial images, giving the system a reasonable invariance to skin-color. These techniques, when combined with the database storage and retrieval mechanisms discussed in the next section, form the basis of a working face recognition system.

LEARNING A FACE

In this implementation, six different sub-windows on the face were extracted. These windows are shown in figure 4-14. For each of the six windows, a gestalt is calculated and transformed for scale. Once the gestalts are calculated for all six windows, all the data is put together as a record in a main database. Included are the filename of the original picture, f-stop, and ID number of the person whose picture it was. (The structure of the main database is shown in Appendix F.) This process is repeated for each picture.

TRAINING FOR AN INDIVIDUAL. Once all desired pictures have been processed, the system is ready to be "trained." The face recognition system characterizes an individual by the X & Y mean and standard deviations of gestalt values over a number of pictures. In this way the system should have an idea of a reasonable range of values to expect for a given individual. For this study, five pictures were taken of each person for training. (It is realized that scores of pictures taken over a period of time (say, a year) would be desirable to thoroughly test the system. However, time constraints prevent this.) It is assumed that five pictures will get us "in the ballpark," and a definable cluster was indeed observed with only 5 pictures.

STATISTICS CALCULATION. Statistics are calculated for

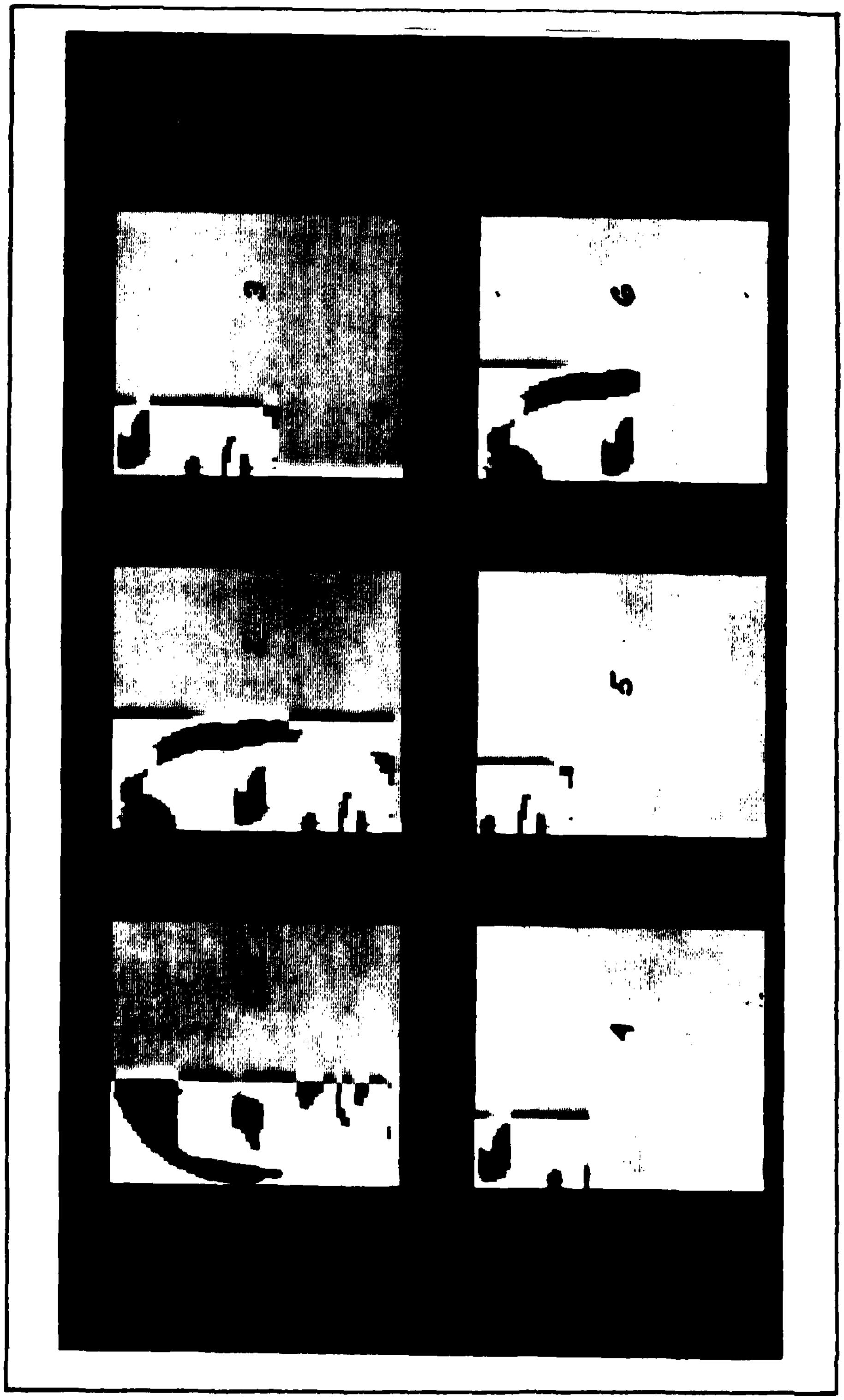


Figure 4-14. Six Windows Used on the Face

each individual in the database, defining their X & Y mean and standard deviations. In addition, overall statistics for each of the windows are calculated, giving such information as how big the search area should be, and which windows give the most reliable information.

RECOGNITION DATABASE GENERATION. A "Recognition Database" is set up for each window, with the ID number and X & Y standard deviations for a person stored at the coordinate location indicated by the person's average gestalt value. Any number of ID numbers can be stored at any coordinate value (see figure 4-15.) All of these values can be retrieved by specifying the coordinate value. For instance, assume that for five pictures, an individual has the following statistics:

X, Y mean = 41,16
X standard deviation = 1.3
Y standard deviation = 2.7
ID number = 1

Therefore, if we accessed the location 41,16 we would find

X standard deviation = 1.3
Y standard deviation = 2.7
ID number = 1

At this point, the coordinate database is completed, and ready to test against for recognition.

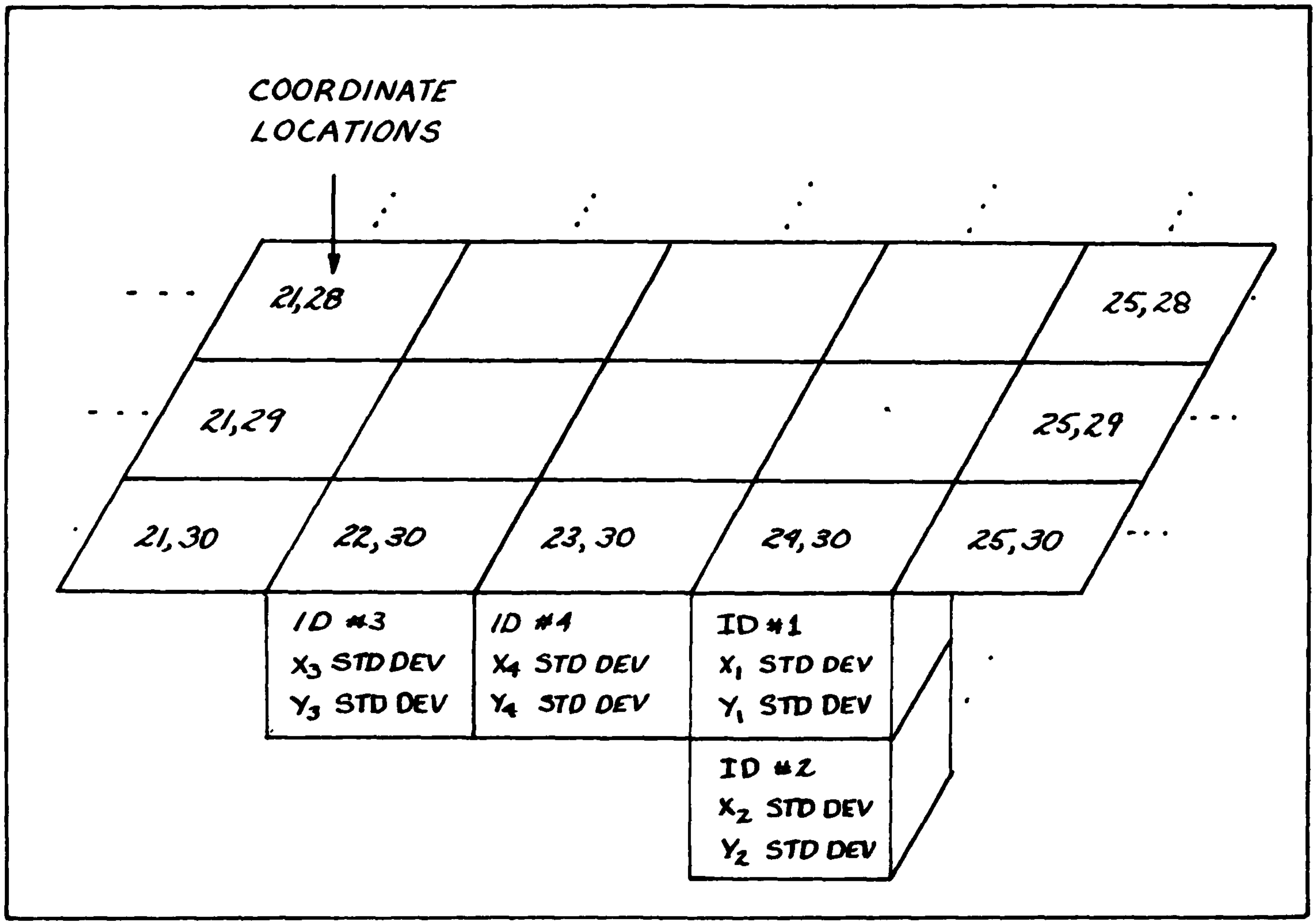


Figure 4-15. Example of Data Storage in a Recognition Database

IDENTIFYING A PERSON

DETERMINATION OF SEARCH SPACE. One of the encouraging results of CTT is that it provides an explanation of the direct memory location phenomenon found in human memory. The human mechanism is able to directly or nearly directly access the particular data regardless of the size of the knowledge base (21). If out of the clear blue someone walks up to a person and says, "Think of your mother-in-law," with seemingly no elapse in time the person can envision her face, feelings he has towards her, what her house looks like, and many other details, even though a moment before he was engrossed in a conversation about the Cincinatti Reds.

The CTT architecture accounts for the direct memory location function by requiring calculation of a 2-dimensional vector as the output of any calculation. The X & Y coordinates of the output then specify the address of the next memory location to be accessed.

This thesis utilizes the direct memory access capability of CTT to restrict the search space required during retrievals, regardless of the size of the knowledge base.

In this concept, the gestalt coordinates for an unidentified person specify the center coordinate value of a search area. then all individuals who have been stored within the search area range are candidates for identification. All others are rejected. This concept is illustrated in figure 4-16.

How should the size of the search area be determined?

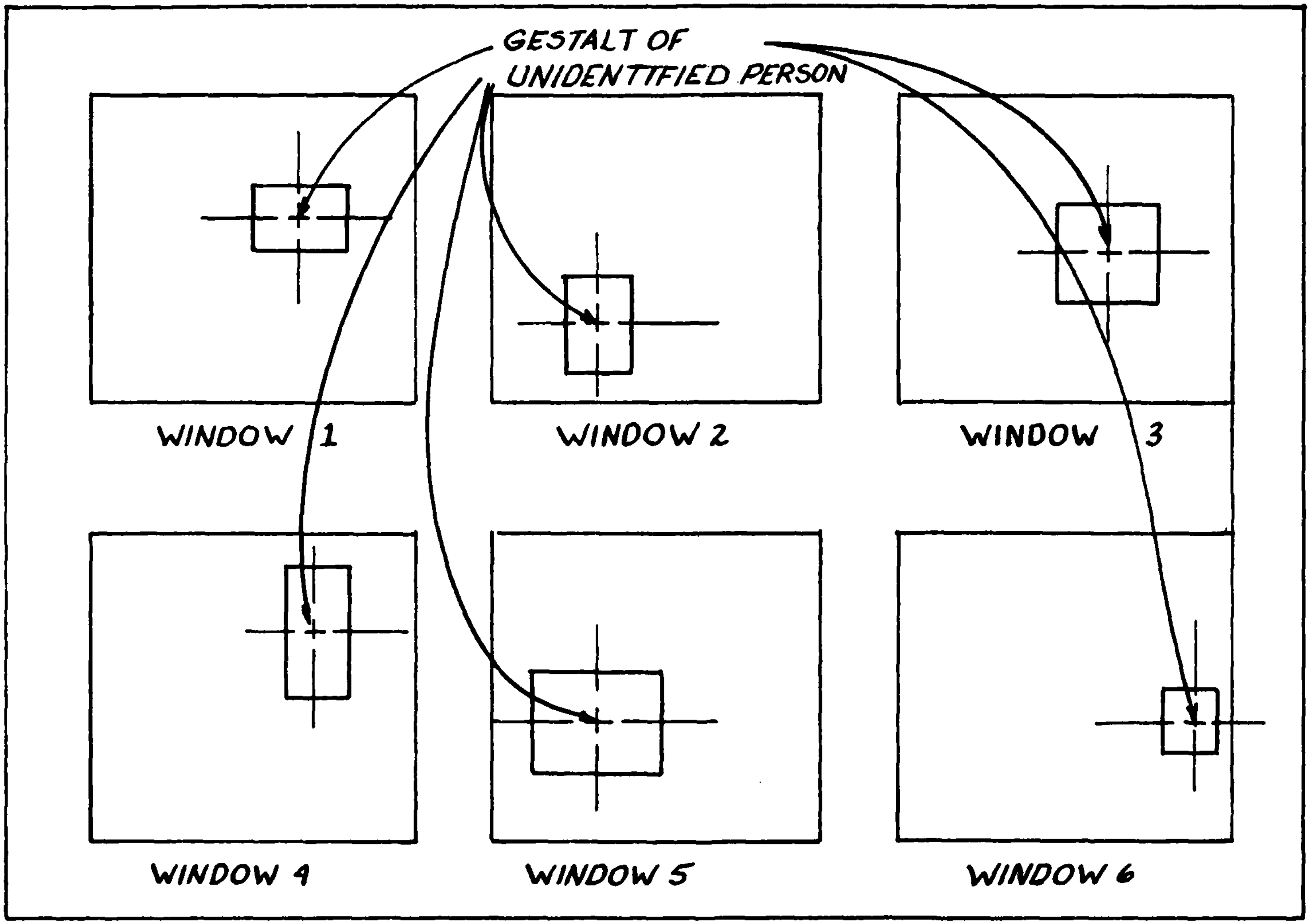


Figure 4-16. Example of Different Search Areas for Different Windows

One method would be to use an arbitrary fixed search area size for all the windows. A problem arises in that the data points for individuals may, on the average, have a bigger spread on one window than another. The method decided upon was to use the average X & Y standard deviations of all individuals in a window as an indicator for how big the search area should be for that window.

For each person in the database, a mean and X & Y standard deviation are calculated from their 2-dimensional gestalt values. This is done for each separate window. Then the average is taken of all the standard deviations of each individual in a particular window for each window. For each window,

$$A_x = \frac{1}{N} \sum_{i=1}^{N} \sigma_{xi} \tag{4-4}$$

and $$A_y = \frac{1}{N} \sum_{i=1}^{N} \sigma_{yi}, \tag{4-5}$$

where σ_x = X Standard Deviation,

σ_y = Y Standard Deviation,

i = Number of particular individual,

N = Number individuals in database

and A_x and A_y are the average standard deviations of the particular window.

In trying to identify an unknown point, the mean coordinate value of the correct individual should be, on the average, within $\pm$ 3 standard deviations of the unknown point

in the X and Y direction (see figure 4-17.) All points that are not within ±3 standard deviations of the unknown point are not considered. This leads to rapid and fairly consistent search times, as a limited subset of the database is all that ever needs to be considered, and the location and size of this area is directly computed (not found through a search technique.)

DESIGN OF DISTANCE METRIC. As discussed in chapter 4, identification in CTT consists of the following steps:

1) Calculate gestalts of several sub-looks (in this case "6") on an unidentified image

2) Each previously-stored individual is represented as a set of 6 gestalt coordinates and standard deviations -- one for each window. Set completion is performed between the set of six gestalts from the unknown individual, and all of the previously-stored individuals. The result is the set of previously-stored gestalts for the individual who matchs most closely to the unidentified points.

3) The gestalt is taken of the 6 coordinate points resulting from set completion. The coordinates of this gestalt give the "name" of the person on the next higher surface.

In CTT, "set completion" is the process which retrieves an entire set of points, given a unique partial set of the points. In this thesis, it is theorized (without proof) that this same process could retrieve a noise-free stored set of points, given a noise-corrupted set of the points

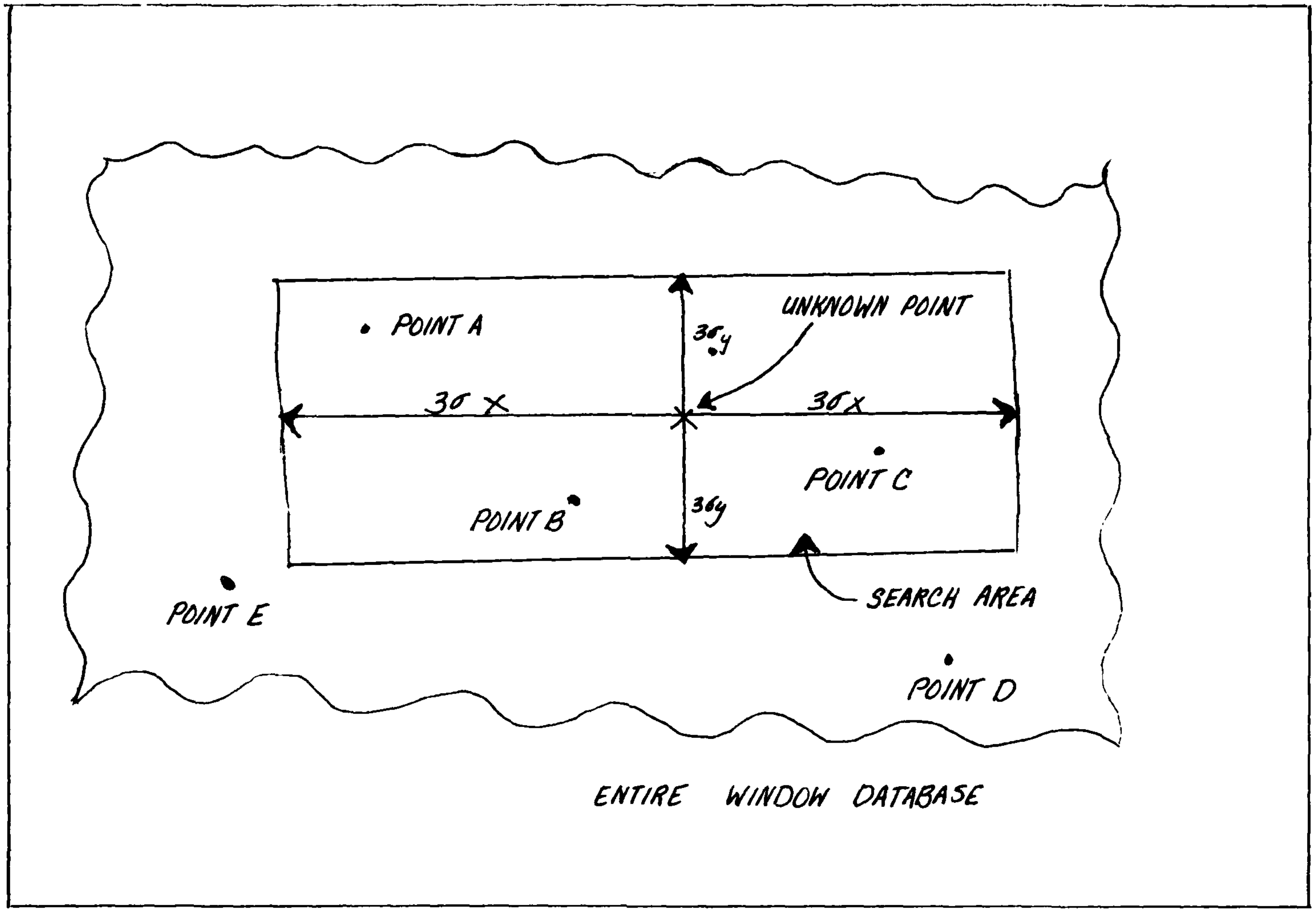


Figure 4-17. Search Area within a Window's Database

which map closer to this set than another set. Such a process could explain many characteristics of the human visual system:

1) How the human visual system can perceive more detail in a picture than is actually there. Harmon's observation of seeing more detail in a discretized facial image, once the image is recognized, is thus explained by set completion, since set completion would "retrieve" the missing details.

2) It explains how what we actually perceive is a function of what we have been conditioned into seeing previously. This is the reason why a trained woodsman can see a squirrel in the woods while his untrained partner might not.

3) It explains how we can look at something, but have problems perceiving it. If we encounter a "new" image which we have not experienced before, there is no previous image to "set complete" with.

4) Since the brain can only process the equivalent of 50 bits of information a second, it is postulated that set completion is necessary to provide the illusion of a higher data processing rate than actually is occuring. It does this by providing extra detail, based on set completion with a partial, lower information image.

It was not an intention of this thesis to try to model the actual implementation of set completion as it occurs (if it occurs) in the cortex, as this is still being researched

and is not yet well understood (6,21). Instead, the overall effect of the process was approximated. First, distances were calculated between mean coordinate values of the unknown points and the known points within the search area of each window. Then distances for each candidate were added for all the windows, and the candidate with the lowest overall distance won.

A reasonable first choice for this operation would be a least-squares fit between the unknown set of gestalts, and each previous set, with the smallest least-square distance winning. This is illustrated as follows:

$$d^2 = \sum_{W=1}^{6} (G_{ix} - G_{ux})^2 + (G_{iy} - G_{uy})^2 \qquad (4\text{-}6)$$

where i = Number of individual stored in database,
W = window number,
G_{ix}, G_{iy} = X, Y gestalt values of stored individual i,
G_{ux}, G_{uy} = X, Y Gestalt values of unidentified individual.

This representation, however, does not take into account that the stored cluster size for one individual may vary from another. Therefore, even though the mean values for two stored individuals may be an equal distance from an unknown point, the individual who has the biggest spread is actually closer. (See figure 4-18.)

This is incorporated in the distance measure as follows:

$$d^2 = \sum_{W=1}^{6} \frac{(G_{ix} - G_{ux})^2}{\sigma_{ix}^2} + \frac{(G_{iy} - G_{uy})^2}{\sigma_{iy}^2} \qquad (4\text{-}7)$$

The CTT gestalt process considers large values to be

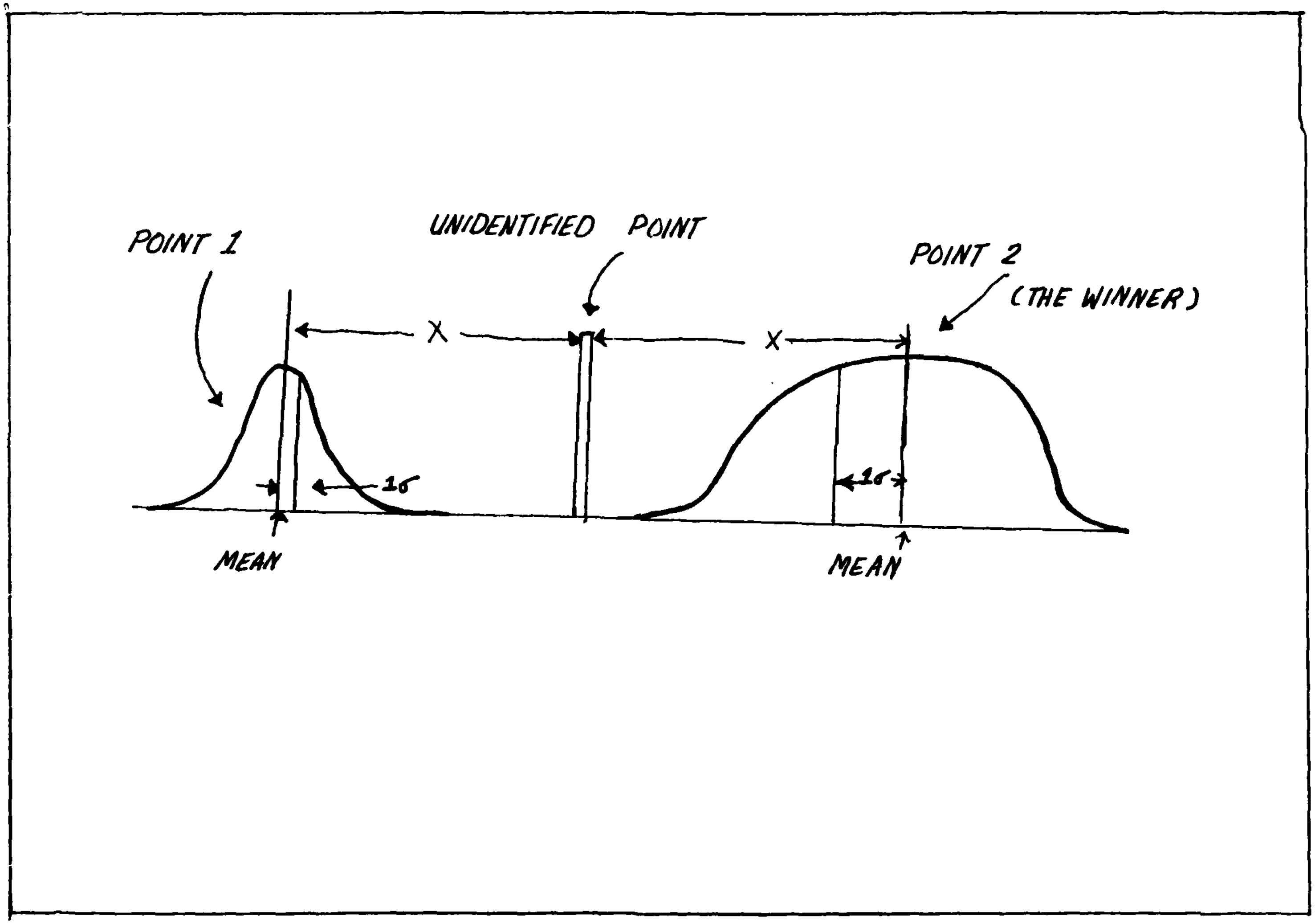


Figure 4-18. Effect of Standard Deviations of Previously-Stored Points on Recognition Decision

more significant than small values. Therefore, to modify this weighting for CTT, the above cost function weighting was changed to the following value function, and weighted by the square root of 2 to "normalize" the function:

$$V_i = \sum_{W=1}^{6} \exp\left\{\frac{-1}{\sqrt{2}}\left[\frac{(G_{ix}-G_{ux})^2}{\sigma_{ix}^2} + \frac{(G_{iy}-G_{uy})^2}{\sigma_{iy}^2}\right]\right\} \tag{4-8}$$

where G_{ix}, G_{iy} = X,Y coordinate values of previously-stored candidate

G_{ux}, G_{uy} = X,Y coordinate values for an unidentified person

and σ_{ix}, σ_{iy} = X,Y standard deviations for person i

This weighting has some nice properties, such as having the value equal 1 when the distance is zero, and having the value decline to zero in a gaussian curve as the distance increases. This also seems to be a reasonable operation for the dendritic network on the cortex to perform, suggesting consistency of this operation with the physiology.

This function was tested for several candidates. A problem was encountered when the unknown point was about 2 standard deviations from the mean value of a stored individual. As previously mentioned, 2 standard deviations away is considered in this thesis to be still within an individual's cluster, and therefore should be given a reasonably high value (see figure 4-19.) However, the above function drops off too quickly, declining to about 30% of max value at 2 standard deviations out. Therefore, the denominator was divided by 2 to spread the funtion out, giving a value of about 90% of max value a 2 standard deviations distance (see figure 4-20.)

The final distance measure for each individual window is:

$$V_i = \exp\left\{\frac{-1}{\sqrt{2}}\left[\frac{(G_{ix}-G_{ux})^2}{(2\sigma_{ix})^2} + \frac{(G_{iy}-G_{uy})^2}{(2\sigma_{iy})^2}\right]\right\} \tag{4-9}$$

where G_{ix}, G_{iy} =X,Y coordinate values of previously stored candidate

G_{ux}, G_{uy} =X,Y coordinate values for an unidentified person

and σ_{ix}, σ_{iy} =X,Y standard deviations for person i

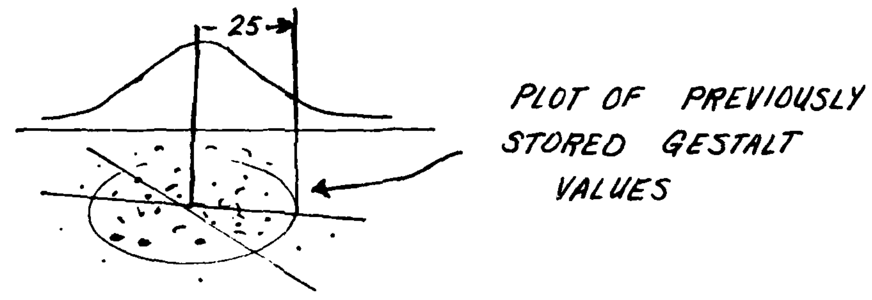


Figure 4-19. Distribution of Stored Gestalt Values

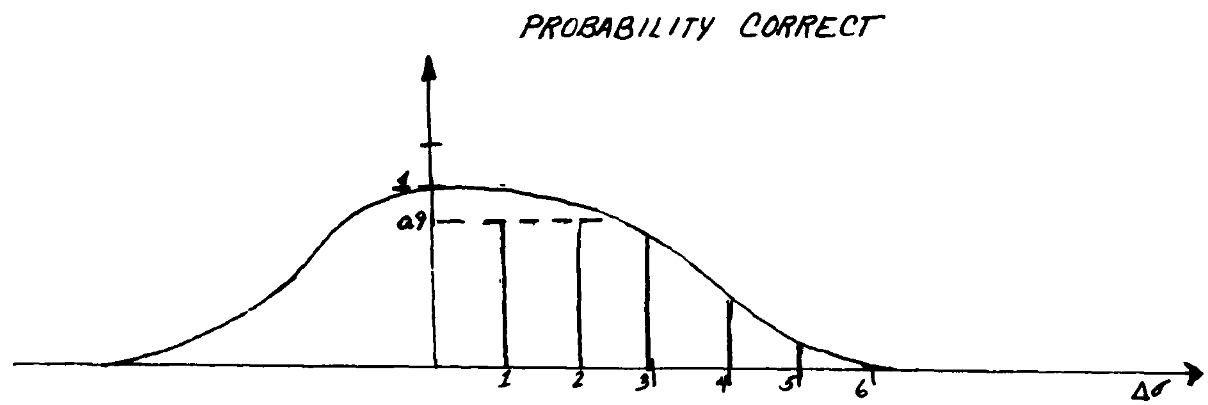


Figure 4-20. Plot of "Probability Correct" vs Magnitude of Standard Deviation Array

COMBINING PROBABILITIES FROM EACH WINDOW

Each window has its own database, and a "probability" value is calculated for each window, or sub-look, on the face. The values from each window are then combined to give the final result, in a manner similar to the "certainty factors" used by MYCIN (20). The probability value from each window represents the strength with which that window suggests similarity to a certain person.

However, when combining values from all the windows, should all windows hold equal weight? Elaine Rich points out that the weighting function should take into account the "confidence in the evidence" (20). In this application, the "confidence" is how well the particular window discriminates between individuals, and is referred to here as "performance factors." Therefore, the final result would be:

V = (Probability of similarity) * (Confidence in window)

for all individuals considered for testing.

The result is a list of candidates by order of overall certainty. How is this confidence for each window measured? A "performance" factor for each window was calculated as follows:

$$P_w = \frac{\text{Average standard deviation of the mean of gestalts}}{\text{Average of the standard deviations for all gestalts}} \tag{4-10}$$

The top term measures how well this window separates the mean values. The bottom term measures how much "spread" there is, on the average, for the individual gestalt values. In general, the performance factor indicates the ability of the window to discriminate between individuals. Figure 4-21 illustrates how this performance rating works. In figure 4-21a, the average standard deviation is small, giving good separation. In figure 4-21b, on the other hand, the average standard deviation is large, even though the mean values have the same separation as in the top figure. As can be seen, the ability to distinguish between the individual elements has gone down, and the performance rating similarly decreases. Now, separating the elements further from each other in figure 4-21c, the standard deviation of the mean values of the elements increases, the ability to distinguish between the elements goes up, and the performance rating similarly increases.

FEATURE MATCHING FUNCTION (INDIVIDUAL WINDOW). Combining window performance factors with the distance calculations for the individual windows, the final value function per window is:

$$V_{iw} = P_w \exp\left\{ \frac{-1}{\sqrt{2}} \left[\frac{(G_{ix} - G_{uy})^2}{(2\sigma_{ix})^2} + \frac{(G_{iy} - G_{uy})^2}{(2\sigma_{iy})^2} \right] \right\} \quad (4\text{-}11)$$

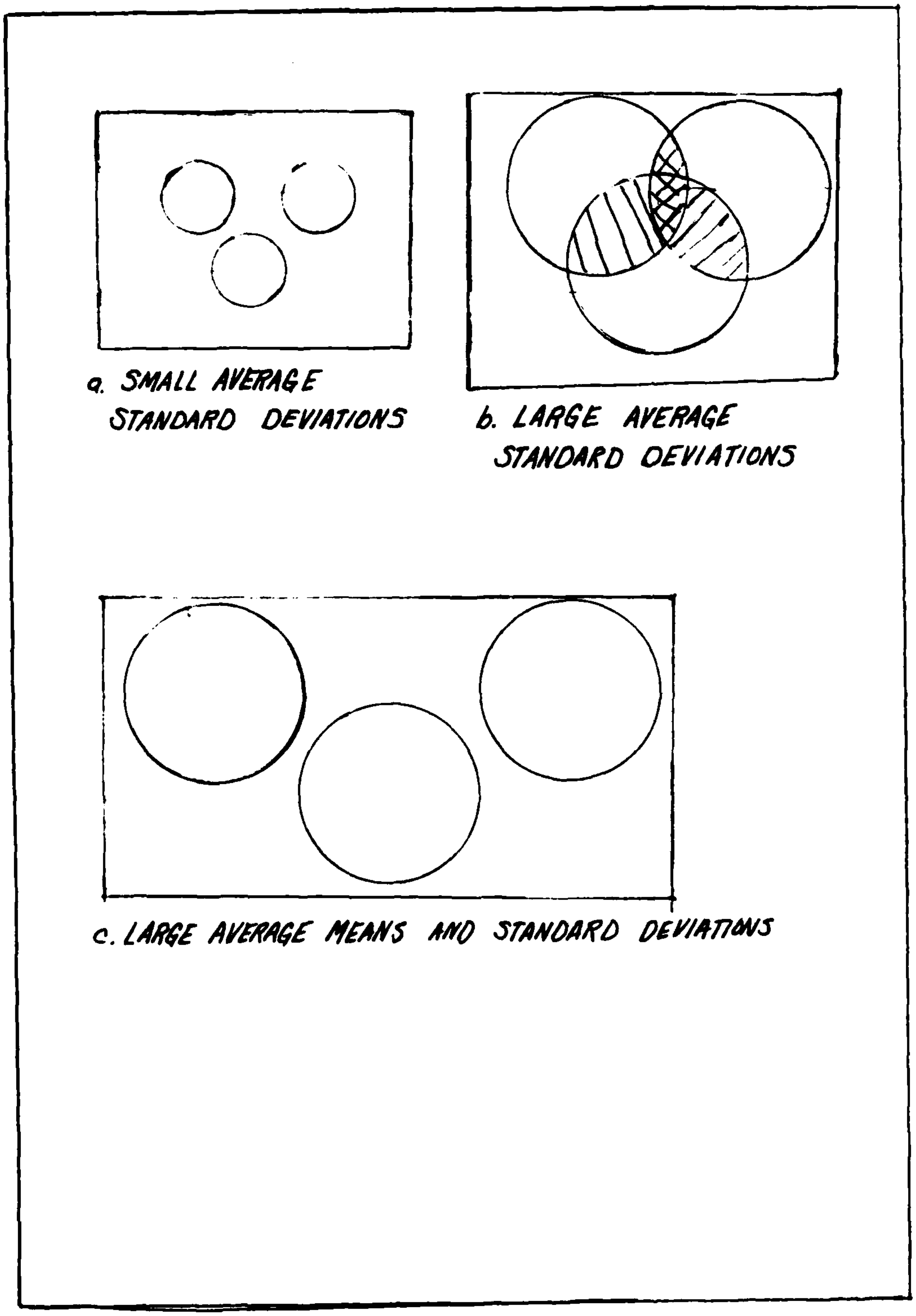


Figure 4-21. Comparison of Separation of Points vs Standard Deviation

where V_{iw} = Value of overall closeness of coordinate values for unidentified person for this window W, with set of coordinate values of previously stored candidate #i for this window

P_w = Performance factor for this window

G_{ix}, G_{iy} = X,Y coordinate values of previously-stored candidate (this window)

G_{ux}, G_{uy} = X,Y coordinate values for an unidentified person (this window)

and σ_{ix}, σ_{iy} = X,Y standard deviations for person i (this window)

This value function is a measure of how closely the set of coordinate points for the unidentified person matches with a previously stored set of points for one of the candidates.

CUMULATIVE FEATURE MATCHING FUNCTION (ALL WINDOWS). By repeating this process for all candidates within the search area of each window, adding the values for each window, and sorting them, the result is a list ordered from the

most-likely candidates to the least-likely.

$$T_i = \sum_{W=1}^{6} V_{iw} \tag{4-12}$$

where T_i = list of total values for individuals for all windows

and V_{iw} = Value of individual #i in window W

SUMMARY

The face recognition system consists of processing of the individual facial images, which can then be used either for training the system for a new individual, or for recognizing an unidentified person. There were at least two key results from this design process:

1) The requirement for displaying facial images as a vertically-split image provides a possible answer to why the human visual system splits its images vertically before displaying them on the left and right primary visual cortexes.

2) Set completion may provide an explanation for several characteristics of the visual system, in cases where the system provides extra detail in an perceived image.

V. IMPLEMENTATION

To test the design ideas, a face recognition system was built at the Signal Processing Lab at the Air Force Institute of Technology. This chapter will discuss the following:

1) Equipment and Studio Setup
2) Steps in the Facial Image Processing
3) General Description of Program Modules
4) Detailed Description of Selected System Components
   a) Image File format
   b) Contrast Enhancement
   c) Feature Location
   d) Calculation of Gestalts, and
   e) Recognition Database
   f) Run Times

The first three sections give the "big picture" of how the system performs its processing. For those interested in more specific implementation details, the "Detailed Description" will discuss key elements of the system.

Equipment and Studio Setup

The following materials and equipment were used:

Data General Eclipse S/250 Computer

Data General Nova 2 Computer

Octek 2000 Video Processing Board

Dage 650 Video Camera with 18-108mm zoom, and f-stop range of 2.5-16

Panasonic WV-5490 Monochrome Monitor

Tektronix 4632 Video Hard Copy Unit

These are configured in a system as shown in figure 5-1. The studio setup is as shown in figure 5-2. The equipment was always arranged in the same location to provide reproducible lighting. The lighting in the Signal Processing Lab consists of overhead flourescent lights aligned parallel along a line from the partition to the camera. The ceiling is 12 feet high. The person to be photographed would sit in front of a partition which had a sheet of white cardboard attached for background.

Steps in Facial Image Processing

Taking the picture. The subject is arranged directly in front of the camera, as shown in figure 5-2. The program PICTURE2.SV is used on the NOVA computer to acquire the picture. The operator adjusts a box-shaped cursor around the head, thereby defining the image to be stored (see figure 5-3.) (The box cursor is adjustable from 64x64 pixels to any smaller values in the X and Y directions.)

Retrieving the Image. Once the initial image has been stored, the picture is then retrieved from disk in a new location on a blank screen, leaving room for further processing steps (see figure 5-4.) (The frequent retrievals from disk, instead of keeping the images in memory, are necessitated by the small memory and program length restrictions of the NOVA computer.)

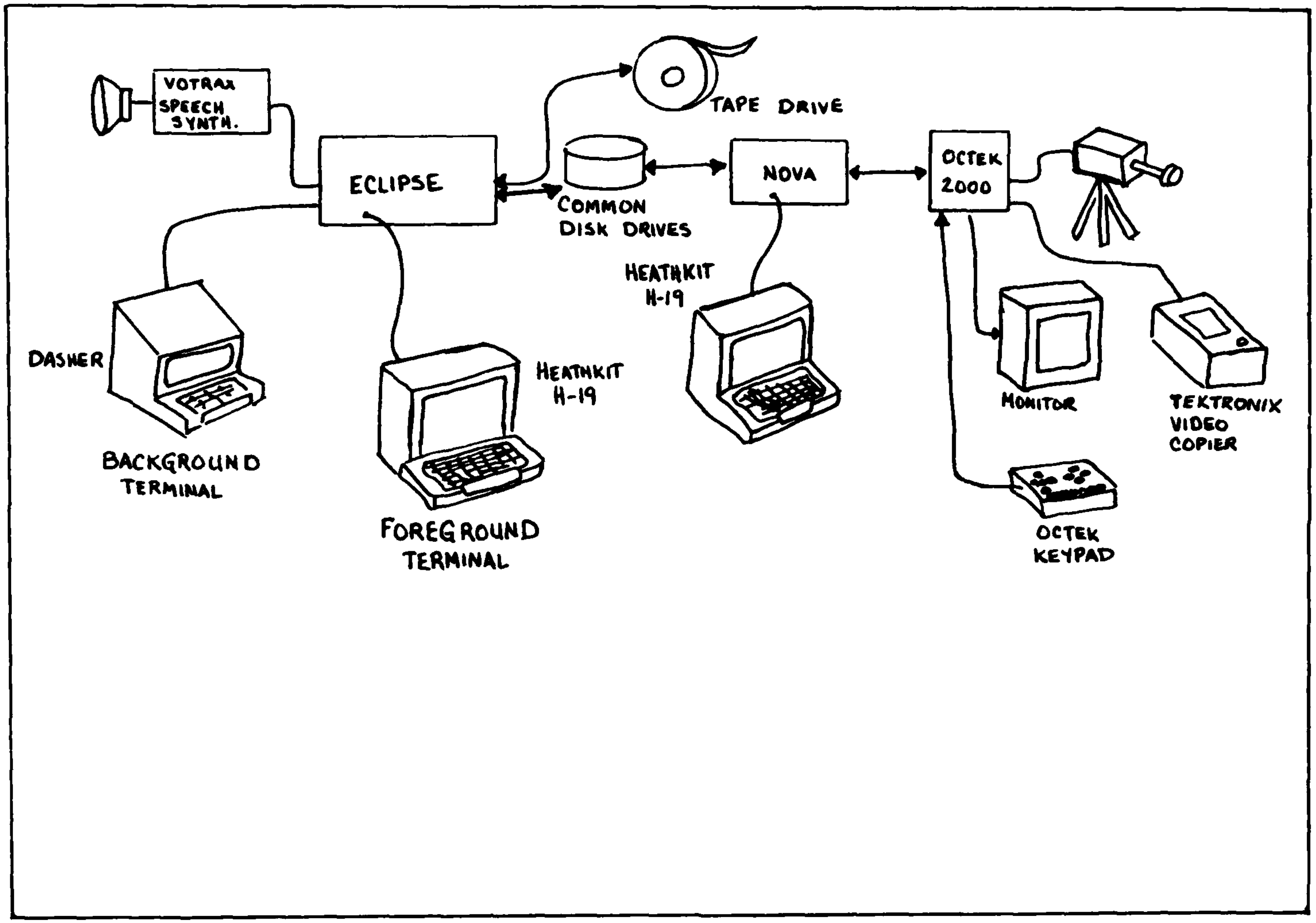


Figure 5-1. System Configuration, AFIT Signal Processing Lab

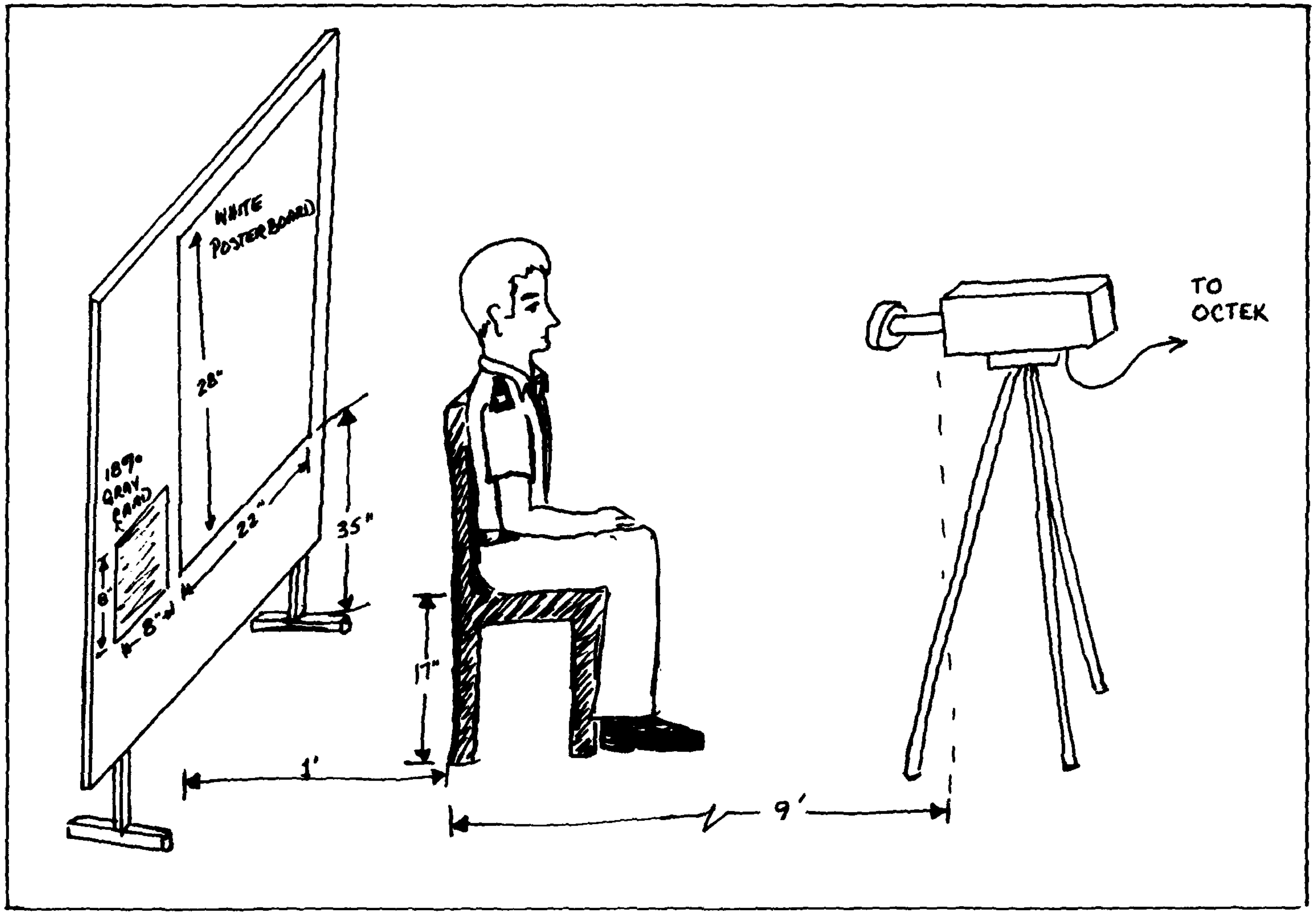


Figure 5-2. Studio Setup for Taking a Picture

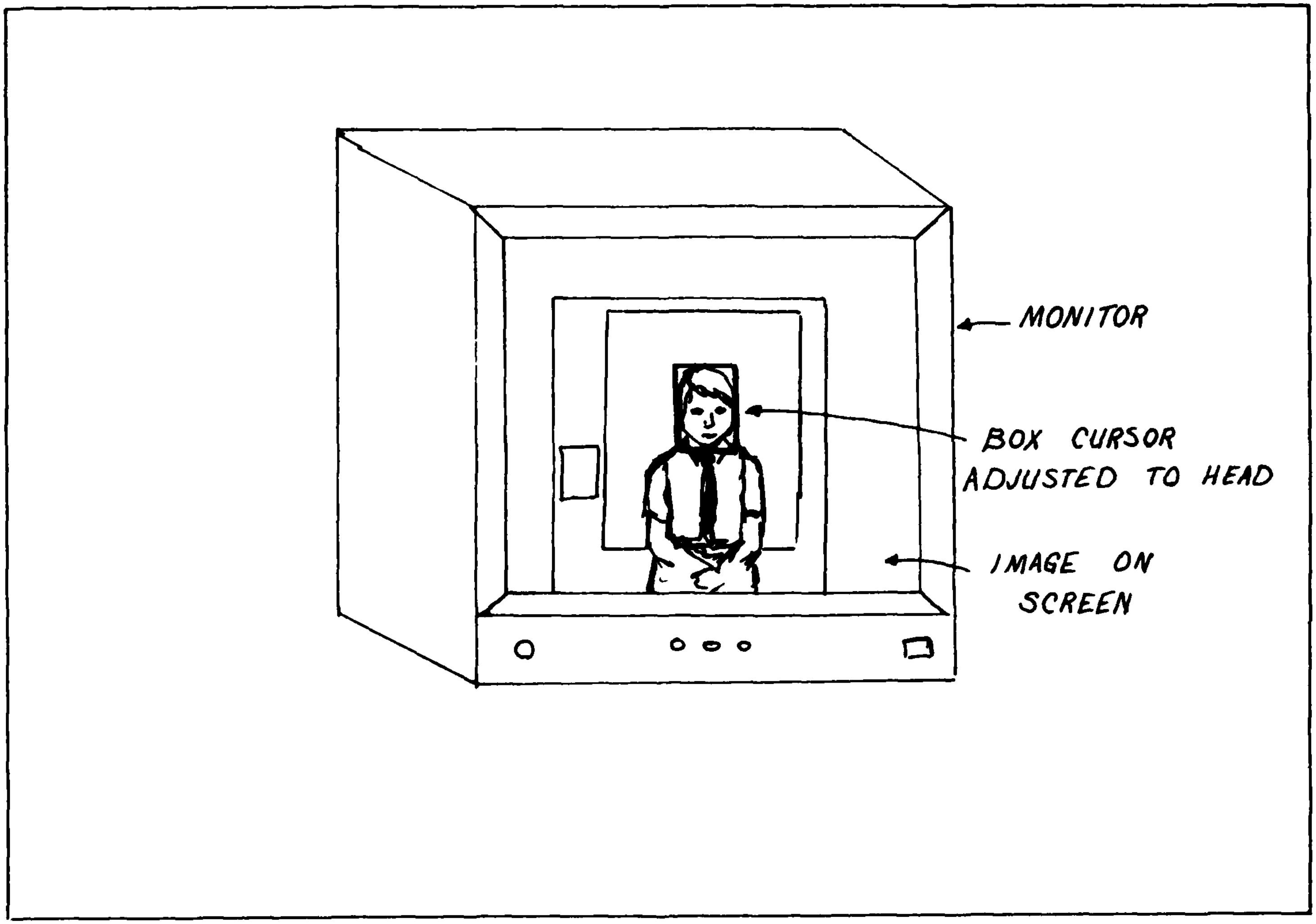


Figure 5-3. Cursor Adjustment on Image

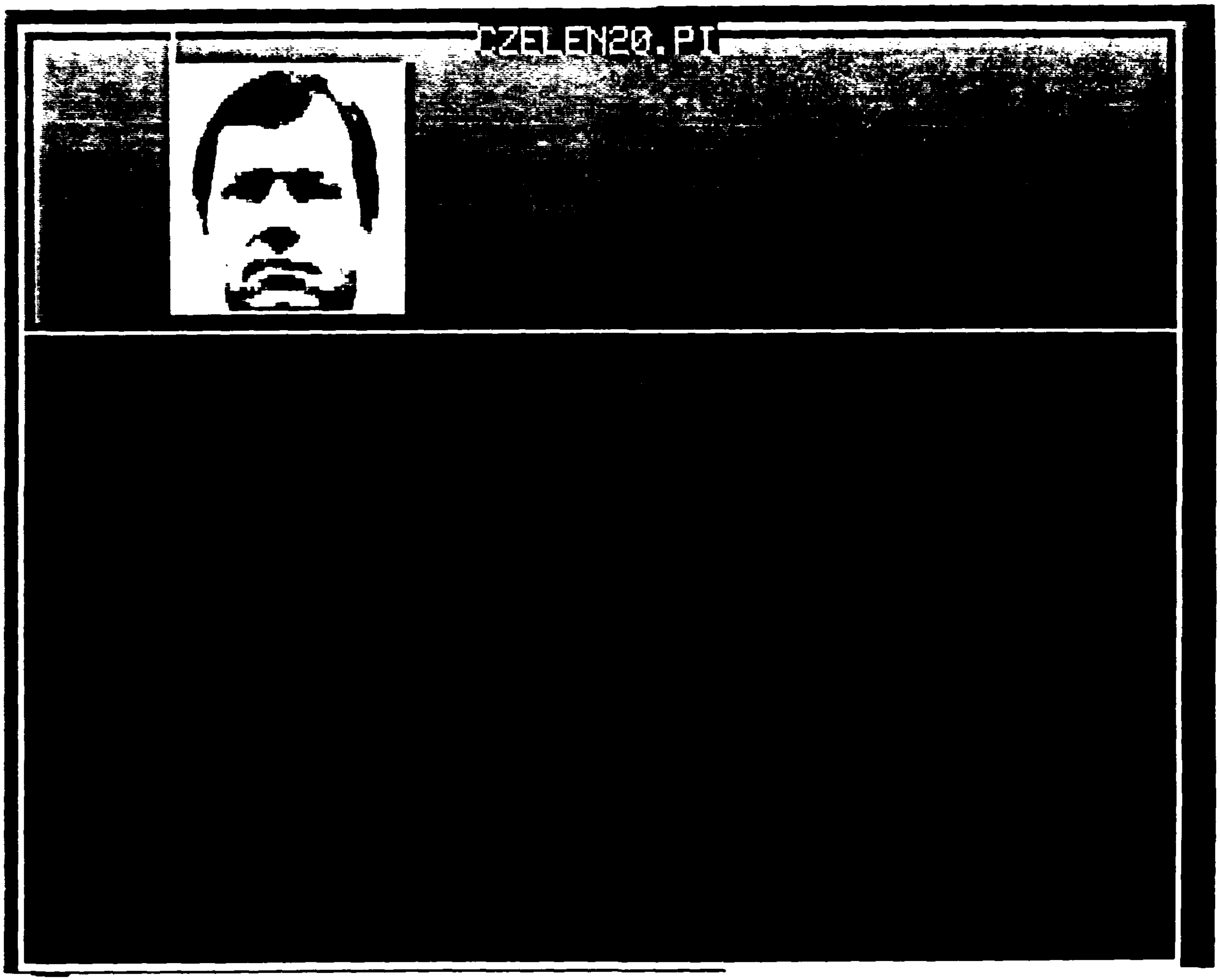


Figure 5-4. Retrieval of Original Image

Contrast Expansion (part 1). The first step of contrast expansion is done on the original image (see chapter 4.) The system measures the average pixel value within a square box cursor centered on the face, as shown in figure 5-5. It then adjusts the contrast based on the average pixel value within the box, by multiplying the value of each pixel in the entire image and limiting the values at the maximum white pixel value (in this case 15) until the average pixel value in the box meets a certain pre-determined value (see figure 5-6.)

Feature Location. Using this contrast-expanded image, the system estimates locations of the major features on the face, and displays them on the screen (see chapter 4 and figure 5-7.) The user can at this point readjust the feature locations if the computer chose them incorrectly. The computer will then redisplay the changed values.

Contrast Expansion (part 2.) As discussed in chapter 4, the system uses the feature locations to do a more precise contrast expansion (see figures 5-8,5-9.) This new image is used for all subsequent processing.

Window Extraction. In this implementation, six different sub-windows on the face were extracted. These windows are shown in figure 5-10. Disk files are created for each one. These images are stored in the same format as the original image. The picture on the screen shows the images displayed on a gray background outside the sub-image,

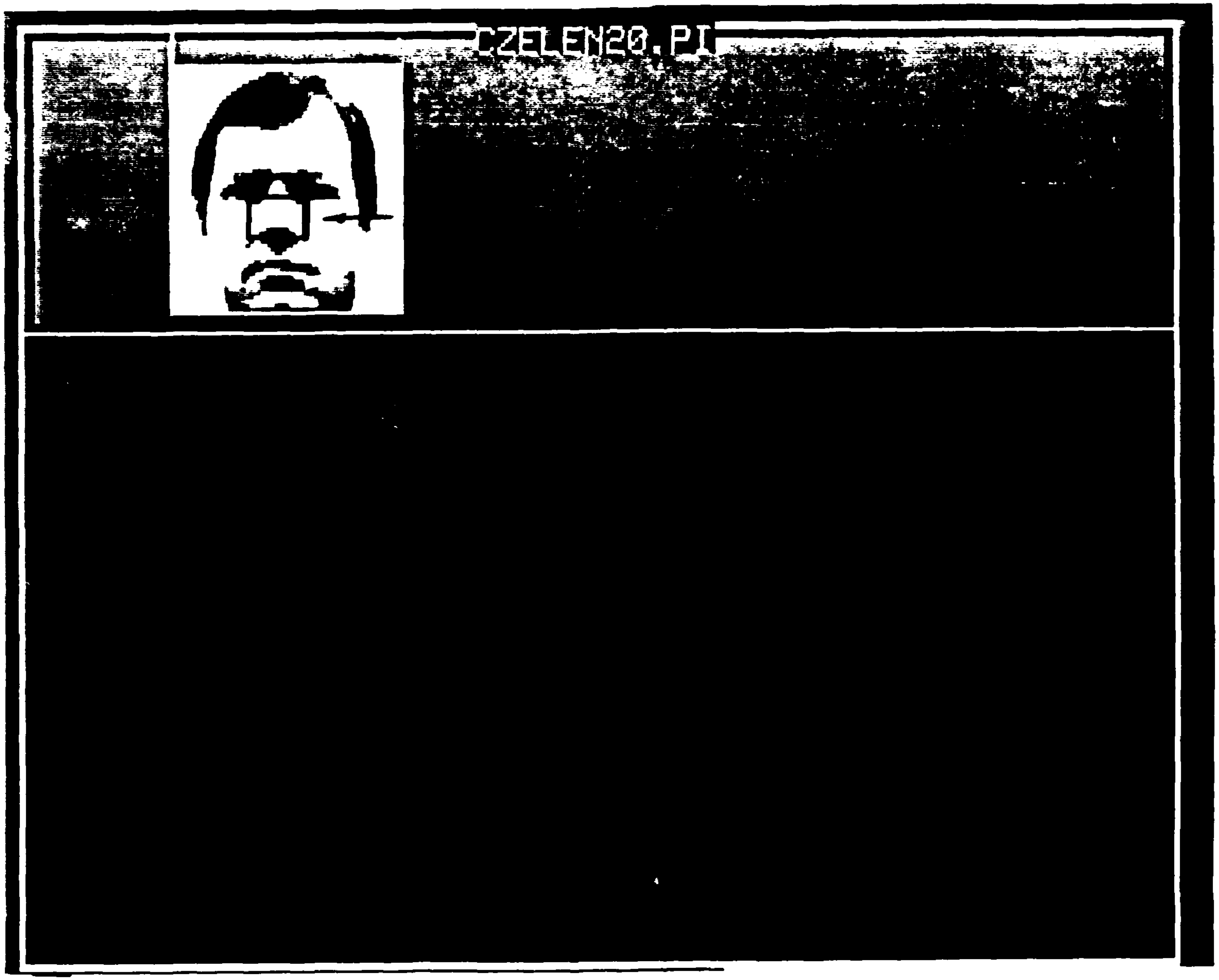


Figure 5-5. Measurement for Initial Contrast Enhancement

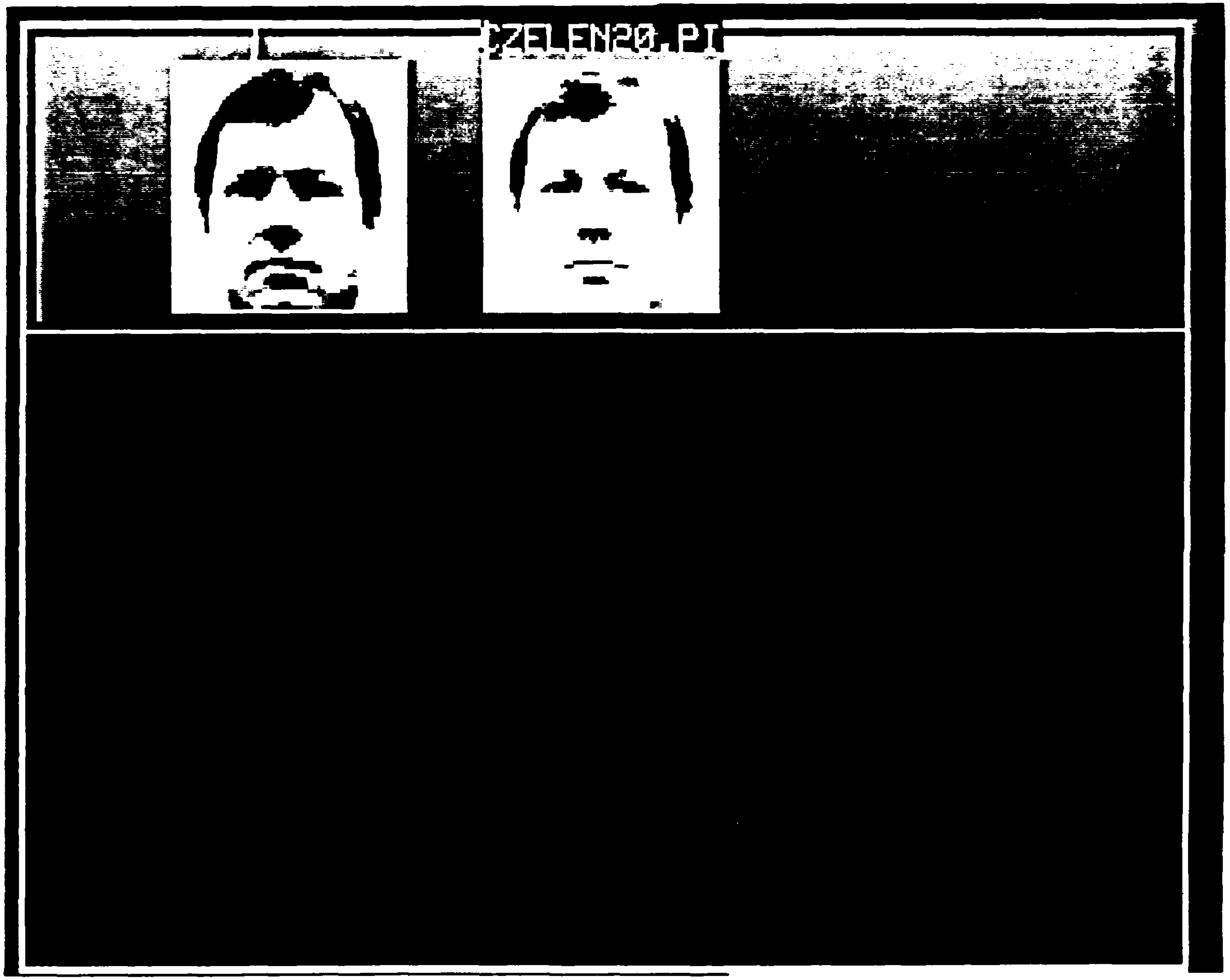


Figure 5-6. Resulting Picture from Initial Contrast Enhancement

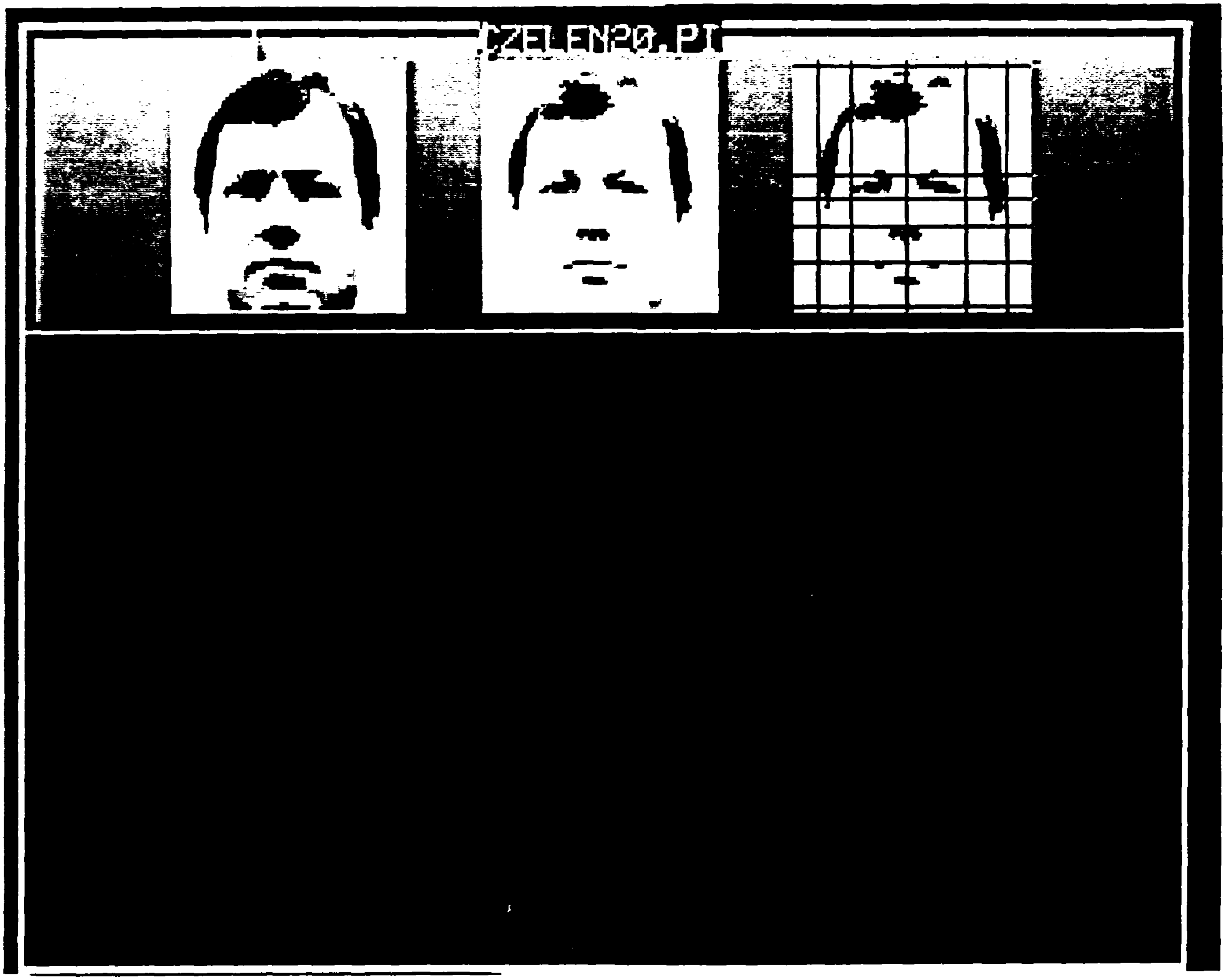


Figure 5-7. Display of Feature Locations

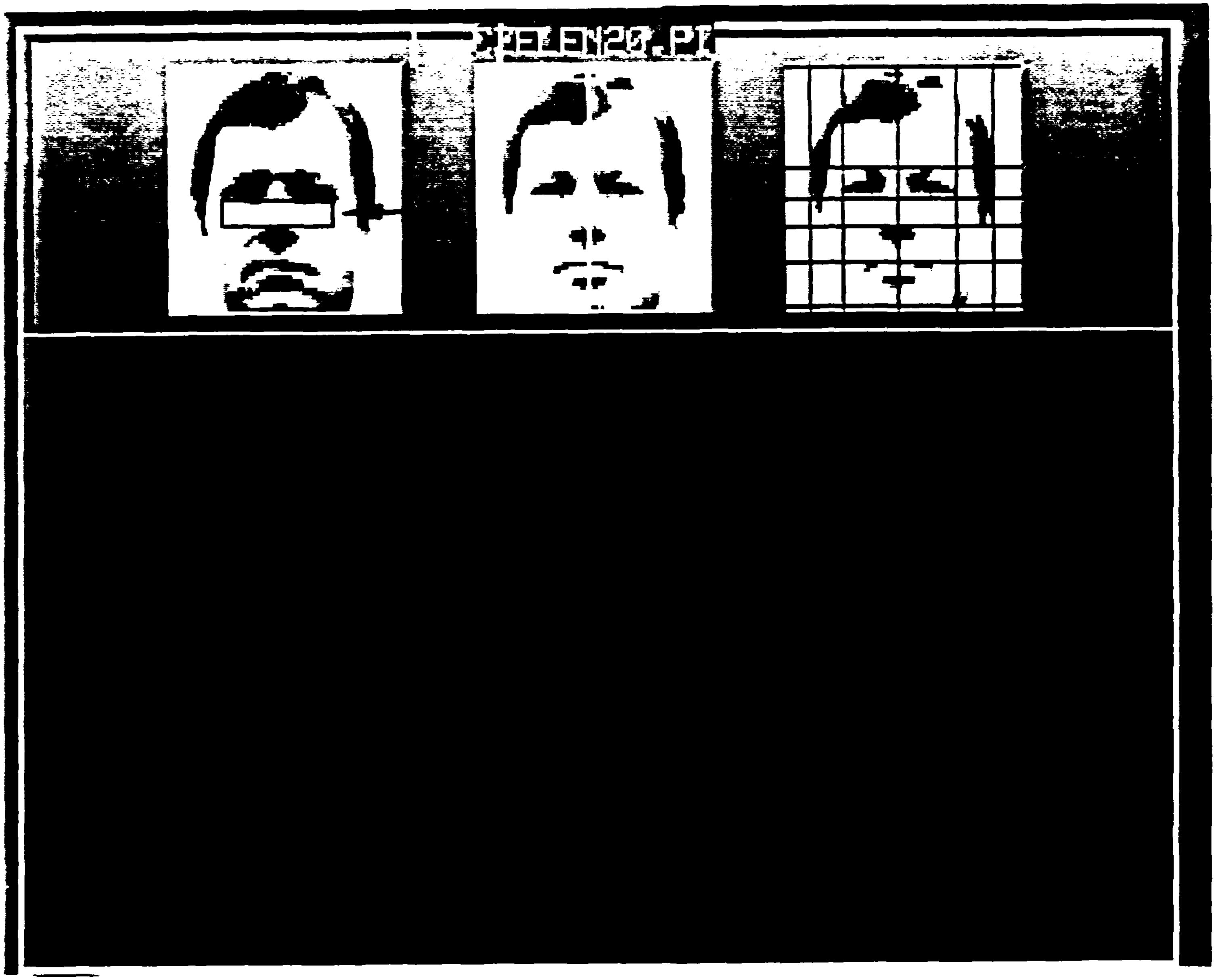


Figure 5-8. Measurement for Final Contrast Enhancement

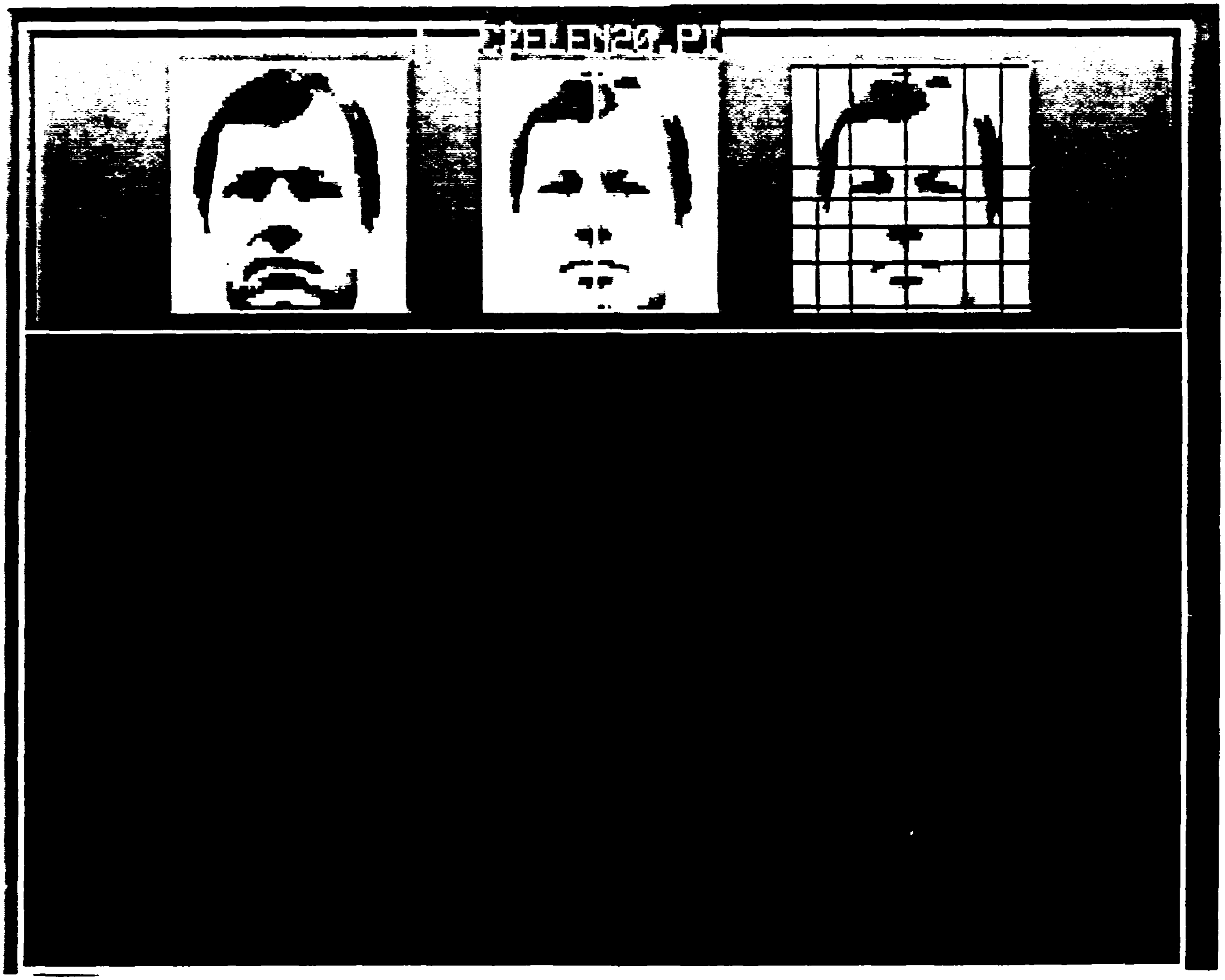


Figure 5-9. Result of Final Contrast Enhancement

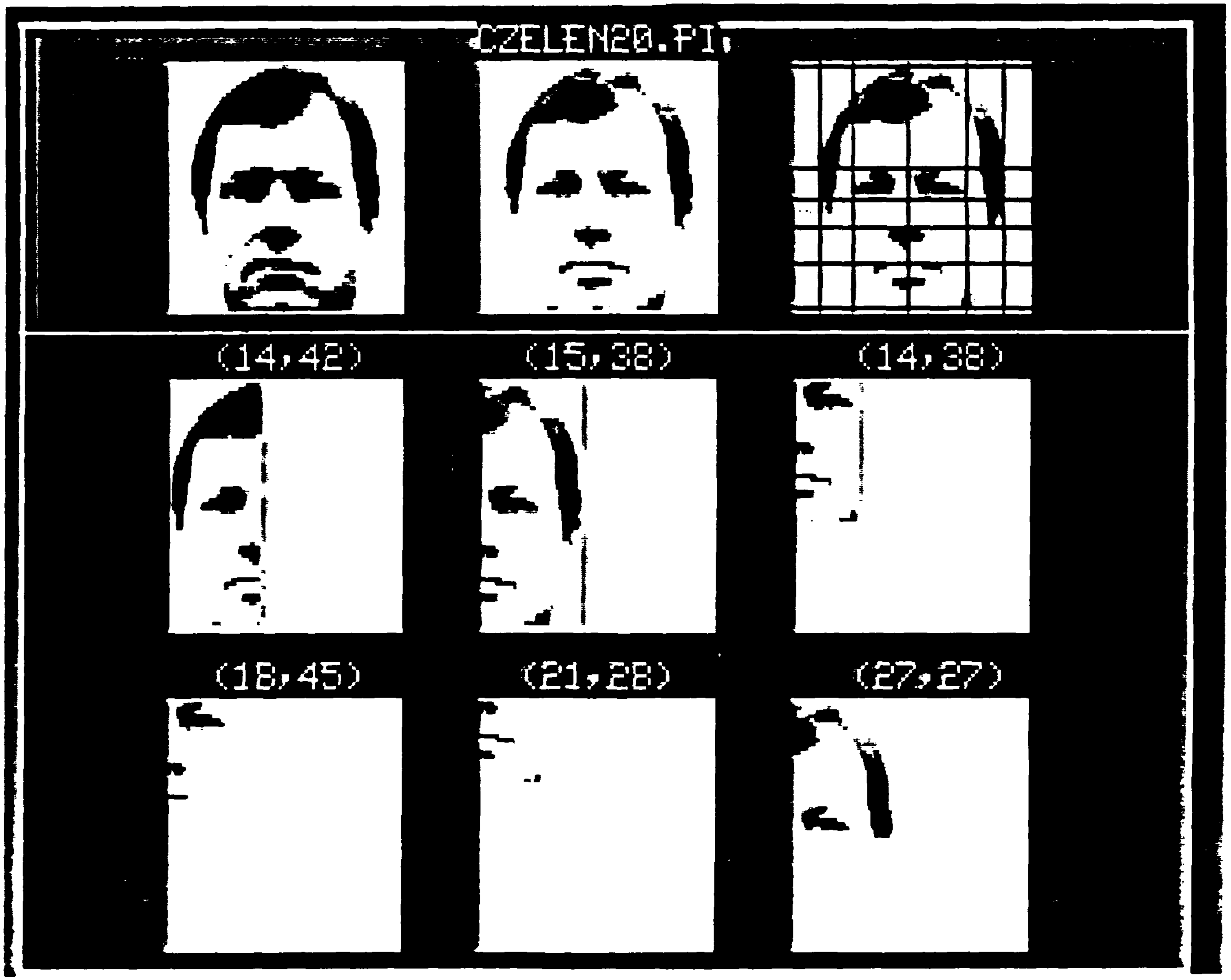


Figure 5-10. Generation of Sub-Images and Display of Gestalts

as opposed to a white background (which is the way the image is stored on disk.) This is done to show the user more clearly the boundaries of the sub-image within the 64x64 box.

Gestalt Calculation. For each of the six windows, a gestalt is calculated (on the Eclipse computer) and transformed for scale (see Table 5-1.) The values are sent back to the Nova, and displayed on the monitor above the sub-image for which it was calculated (see figure 5-10.)

Storage in Database. Once the gestalts are calculated for all six windows on the face, all the data is put together as a record in the Processed Picture Database (the file called "MAINPICS".) Included are the filename of the original picture, f-stop, and ID number of the person whose picture it was, along with other data. (The structure of records in the Processed Picture Database is shown in Appendix F.)

This process is repeated for each picture. When all the pictures are entered for an individual, the system is ready to be "trained" with the data.

Training the Database. Once all desired pictures have been processed, the system is ready to be "trained." This is done by selecting "Calculate Statistics" on the main program running on the Eclipse computer. The face recognition system characterizes an individual by the X & Y mean and standard deviations of gestalt values over a number

```
          * * * Gestalt Calculations for DOYLE * * *

Date:  11/07/85     Time:  15:47

* * * Coordinate Points for WINDOW #1 * * *
   X, Y WINDOW SIZE = 26,61
   ORIGINAL GESTALT = (15,40)
   GESTALT (Windowed by X & Y) = (37,39)
   GESTALT (Windowed by max of X & Y) = (16,39) <<== Final
   AMPLITUDE =      3362                             Answer

* * * Coordinate Points for WINDOW #2 * * *
   X, Y WINDOW SIZE = 27,61
   ORIGINAL GESTALT = (12,36)
   GESTALT (Windowed by X & Y) = (28,35)
   GESTALT (Windowed by max of X & Y) = (13,35) <<== Final
   AMPLITUDE =      3155                             Answer

* * * Coordinate Points for WINDOW #3 * * *
   X, Y WINDOW SIZE = 14,39
   ORIGINAL GESTALT = ( 7,45)
   GESTALT (Windowed by X & Y) = (32,32)
   GESTALT (Windowed by max of X & Y) = (11,32) <<== Final
   AMPLITUDE =      1593                             Answer

* * * Coordinate Points for WINDOW #4 * * *
   X, Y WINDOW SIZE = 14,29
   ORIGINAL GESTALT = ( 7,49)
   GESTALT (Windowed by X & Y) = (32,30)
   GESTALT (Windowed by max of X & Y) = (15,30) <<== Final
   AMPLITUDE =      1308                             Answer

* * * Coordinate Points for WINDOW #5 * * *
   X, Y WINDOW SIZE = 14,21
   ORIGINAL GESTALT = ( 6,55)
   GESTALT (Windowed by X & Y) = (27,35)
   GESTALT (Windowed by max of X & Y) = (18,35) <<== Final
   AMPLITUDE =      1051                             Answer

* * * Coordinate Points for WINDOW #6 * * *
   X, Y WINDOW SIZE = 27,34
   ORIGINAL GESTALT = (13,48)
   GESTALT (Windowed by X & Y) = (31,33)
   GESTALT (Windowed by max of X & Y) = (24,33) <<== Final
   AMPLITUDE =      2303                             Answer
```

Table 5-1. Example Output of Gestalt Calculations

of pictures. In this way the system should have an idea of a reasonable range of values to expect for a given individual. For this study, five pictures were taken of each person for training. The author realized that scores of pictures taken over a period of time (say, a year) would be desirable to thoroughly test the system. However, time constraints prevent this. It was assumed the five pictures would get us "in the ballpark," and a definable cluster was indeed observed with only 5 pictures.

Statistics were calculated for each individual in the database, defining his X & Y mean and standard deviations. In addition, overall statistics for each of the windows were calculated, giving such information as how big the search area should be, and which windows give the most reliable information. A "Recognition Database" was set up for each of the six windows, with the ID number and X & Y standard deviations for a person stored at the coordinate location indicated by the person's average gestalt value. (See chapter 4 for specifics on the database design.) In addition, these values were retrieved by specifying the coordinate value. For instance, assume that for five pictures, an individual has the following statistics in window 1 (left side of face):

X, Y mean = 41,16

X standard deviation = 1.3

* * * FACE RECOGNITION DATABASE -- STATISTICS CALCULATIONS * * *

(The smallest standard deviation is defined to be 0.5, in order to take care of discretation error.)

Date: 11/27/85   Time: 17:28

* * * CALCULATIONS FOR WINDOW 1 * * *

* * * STATISTICS FOR ID NUMBER   1, CAPT RON SMALL   * * *

Total Number of Points in Database =   8
X Standard Deviation =   .60
Y Standard Deviation =   1.66
Average X Value =   12.1
Average Y Value =   46.0

Minimum X Distance = 11
Maximum X Distance = 13
Minimum Y Distance = 44
Maximum Y Distance = 48

* * * STATISTICS FOR ID NUMBER   2, CAPT BOB RUSSEL   * * *

Total Number of Points in Database =   9
X Standard Deviation =   .67
Y Standard Deviation =   .50
Average X Value =   14.0
Average Y Value =   47.0

Minimum X Distance = 13
Maximum X Distance = 15
Minimum Y Distance = 46
Maximum Y Distance = 48

* * * STATISTICS FOR ID NUMBER   3, CAPT MAX HALL   * * *

Total Number of Points in Database =   10
X Standard Deviation =   .50
Y Standard Deviation =   .92
Average X Value =   14.0
Average Y Value =   40.4

Minimum X Distance = 14
Maximum X Distance = 14

Table 5-2. Example of Statistics Calculations (for 1 window)

Minimum Y Distance = 39
Maximum Y Distance = 42

* * * STATISTICS FOR ID NUMBER 4, CAPT JERRY GERACE * * *

Total Number of Points in Database = 9
X Standard Deviation = .63
Y Standard Deviation = .69
Average X Value = 13.2
Average Y Value = 45.4

Minimum X Distance = 12
Maximum X Distance = 14
Minimum Y Distance = 44
Maximum Y Distance = 46

* * * STATISTICS FOR ID NUMBER 5, CAPT TOM GRIFFIN * * *

Total Number of Points in Database = 10
X Standard Deviation = .70
Y Standard Deviation = 1.36
Average X Value = 12.9
Average Y Value = 31.6

Minimum X Distance = 12
Maximum X Distance = 14
Minimum Y Distance = 29
Maximum Y Distance = 33

* * * STATISTICS FOR ID NUMBER 6, DR TERRY SKELTON * * *

Total Number of Points in Database = 9
X Standard Deviation = .74
Y Standard Deviation = 1.29
Average X Value = 16.1
Average Y Value = 39.1

Minimum X Distance = 15
Maximum X Distance = 17
Minimum Y Distance = 37
Maximum Y Distance = 41

* * * STATISTICS FOR ID NUMBER 7, CAPT DAVE HUNSUCK * * *

****

* * * SUMMARY OF WINDOW PERFORMANCES * * *

Window Number	X Perf.	Y Perf.	Figure of Merit
1	2.13	8.41	8.67
2	3.20	7.53	8.18
3	1.91	3.88	4.32
4	1.80	3.82	4.22
5	2.13	2.41	3.22
6	3.40	7.78	8.49

****

```
* * * CTT FACE RECOGNITION SYSTEM * * *
Date: 11/27/85    Time: 17:33


Filename of picture being recognized = SMALL9.PI



   * * * CANDIDATES FOR WINDOW 1 * * *

X,Y Location of Unidentified Person: 12,43    X Sigma (for Window) =    .67
    Y Sigma (for Window) =     .95
Number of Sigmas Out We're Searching =   3.0
Range of Search:  X Coordinate = 10 to 14.  Y Coordinate = 40 to 46.

ID Number =   16   CAPT JIM HOLTEN          Position = 13,40  Prob =    5.17
        Sigmas Away -- X:    1.58    Sigmas Away -- Y:    1.28
ID Number =   13   CAPT PHIL FITZJARREL     Position = 13,40  Prob =     .06
        Sigmas Away -- X:    2.00    Sigmas Away -- Y:    6.00
ID Number =    3   CAPT MAX HALL            Position = 14,40  Prob =     .30
        Sigmas Away -- X:    4.00    Sigmas Away -- Y:    3.29
ID Number =   10   MR. SWAMI KRISHNASWAMI   Position = 11,44  Prob =    6.66
        Sigmas Away -- X:    1.25    Sigmas Away -- Y:     .74
ID Number =    4   CAPT JERRY GERACE        Position = 13,45  Prob =    2.13
        Sigmas Away -- X:    1.61    Sigmas Away -- Y:    2.94
ID Number =    1   CAPT RON SMALL           Position = 12,46  Prob =    5.76
        Sigmas Away -- X:     .00    Sigmas Away -- Y:    1.81
ID Number =   11   CAPT FRED STIERWALT      Position = 14,46  Prob =     .12
        Sigmas Away -- X:    4.00    Sigmas Away -- Y:    4.28



   * * * CANDIDATES FOR WINDOW 2 * * *

X,Y Location of Unidentified Person: 10,43    X Sigma (for Window) =    .87
    Y Sigma (for Window) =    1.00
Number of Sigmas Out We're Searching =   3.0
Range of Search:  X Coordinate =  7 to 13.  Y Coordinate = 40 to 46.

ID Number =   11   CAPT FRED STIERWALT      Position = 10,40  Prob =    3.29
        Sigmas Away -- X:     .00    Sigmas Away -- Y:    2.70
ID Number =    4   CAPT JERRY GERACE        Position = 11,43  Prob =    6.83
        Sigmas Away -- X:    1.20    Sigmas Away -- Y:     .00
ID Number =   18   CAPT RIC ROUTH           Position = 11,45  Prob =    2.65
        Sigmas Away -- X:     .95    Sigmas Away -- Y:    2.85
ID Number =    1   CAPT RON SMALL           Position = 11,45  Prob =    5.79
        Sigmas Away -- X:    1.17    Sigmas Away -- Y:    1.18
```

Table 5-3. Output from Recognition Process

```
ID Number =    2   CAPT BOB RUSSEL           Position = 10,46   Prob =     1.86
       Sigmas Away -- X:      .00     Sigmas Away -- Y:     3.44



    * * * CANDIDATES FOR WINDOW 3 * * *

X,Y Location of Unidentified Person:  9,34    X Sigma (for Window) =      .81
    Y Sigma (for Window) =     1.93
Number of Sigmas Out We're Searching =    3.0
Range of Search:  X Coordinate =  7 to 11.  Y Coordinate = 28 to 40.

ID Number =    7   CAPT DAVE HUNSUCK         Position = 11,35   Prob =      .57
       Sigmas Away -- X:     4.00     Sigmas Away -- Y:      .46
ID Number =    1   CAPT RON SMALL            Position = 10,36   Prob =     3.18
       Sigmas Away -- X:      .90     Sigmas Away -- Y:     1.28



    * * * CANDIDATES FOR WINDOW 4 * * *

X,Y Location of Unidentified Person: 12,33    X Sigma (for Window) =     1.04
    Y Sigma (for Window) =     1.49
Number of Sigmas Out We're Searching =    3.0
Range of Search:  X Coordinate =  9 to 15.  Y Coordinate = 29 to 37.

ID Number =    1   CAPT RON SMALL            Position = 14,31   Prob =     2.44
       Sigmas Away -- X:     1.85     Sigmas Away -- Y:      .98



    * * * CANDIDATES FOR WINDOW 5 * * *

X,Y Location of Unidentified Person: 16,33    X Sigma (for Window) =     1.92
    Y Sigma (for Window) =     2.30
Number of Sigmas Out We're Searching =    3.0
Range of Search:  X Coordinate = 10 to 22.  Y Coordinate = 26 to 40.

ID Number =   13   CAPT PHIL FITZJARREL      Position = 19,26   Prob =      .00
       Sigmas Away -- X:     2.04     Sigmas Away -- Y:     8.53
ID Number =    8   CAPT MARK CLIFFORD        Position = 17,27   Prob =     1.33
       Sigmas Away -- X:      .29     Sigmas Away -- Y:     2.64
ID Number =    9   DR. WOODROW W. BLEDSOE    Position = 20,27   Prob =      .03
       Sigmas Away -- X:     4.93     Sigmas Away -- Y:     3.68
ID Number =   10   MR. SWAMI KRISHNASWAMI    Position = 18,29   Prob =     2.81
       Sigmas Away -- X:      .70     Sigmas Away -- Y:      .77
```

```
ID Number =   12   CAPT MIKE HUNSUCKER        Position = 21,29   Prob =     .49
       Sigmas Away -- X:     3.16     Sigmas Away -- Y:     2.24
ID Number =   18   CAPT RIC ROUTH             Position = 20,30   Prob =    1.60
       Sigmas Away -- X:     1.52     Sigmas Away -- Y:     1.81
ID Number =    4   CAPT JERRY GERACE          Position = 16,31   Prob =    3.05
       Sigmas Away -- X:      .00     Sigmas Away -- Y:      .65
ID Number =   17   DR. BILL CZELEN            Position = 17,31   Prob =    1.84
       Sigmas Away -- X:     2.00     Sigmas Away -- Y:      .69
ID Number =   16   CAPT JIM HOLTEN            Position = 20,31   Prob =    1.93
       Sigmas Away -- X:     1.91     Sigmas Away -- Y:      .66
ID Number =   20   MRS. EDIE ROUTH            Position = 21,31   Prob =    2.13
       Sigmas Away -- X:     1.51     Sigmas Away -- Y:     1.00
ID Number =    7   CAPT DAVE HUNSUCK          Position = 16,32   Prob =    3.16
       Sigmas Away -- X:      .00     Sigmas Away -- Y:      .39
ID Number =    3   CAPT MAX HALL              Position = 17,32   Prob =    3.08
       Sigmas Away -- X:      .44     Sigmas Away -- Y:      .40
ID Number =    5   CAPT TOM GRIFFIN           Position = 11,37   Prob =    2.01
       Sigmas Away -- X:     1.66     Sigmas Away -- Y:     1.00
ID Number =    2   CAPT BOB RUSSEL            Position = 11,37   Prob =     .31
       Sigmas Away -- X:     3.59     Sigmas Away -- Y:     2.40
ID Number =   14   CAPT DAVID KING            Position = 10,38   Prob =     .33
       Sigmas Away -- X:     3.68     Sigmas Away -- Y:     2.18
ID Number =    6   DR TERRY SKELTON           Position = 12,39   Prob =     .50
       Sigmas Away -- X:     2.87     Sigmas Away -- Y:     2.59
ID Number =    1   CAPT RON SMALL             Position = 14,40   Prob =    1.58
       Sigmas Away -- X:      .78     Sigmas Away -- Y:     2.25


    * * * CANDIDATES FOR WINDOW 6 * * *

X,Y Location of Unidentified Person: 20,32    X Sigma (for Window) =    1.50
    Y Sigma (for Window) =    1.01
Number of Sigmas Out We're Searching =   3.0
Range of Search:  X Coordinate = 16 to 24.  Y Coordinate = 29 to 35.

ID Number =   11   CAPT FRED STIERWALT        Position = 18,30   Prob =     .85
       Sigmas Away -- X:     1.55     Sigmas Away -- Y:     4.00
ID Number =    4   CAPT JERRY GERACE          Position = 19,33   Prob =    7.20
       Sigmas Away -- X:      .95     Sigmas Away -- Y:      .65
ID Number =    1   CAPT RON SMALL             Position = 21,35   Prob =    3.09
       Sigmas Away -- X:      .65     Sigmas Away -- Y:     2.77


    * * * COMPUTER'S CHOICE(S) FOR WHO THIS IS * * *

ID Number =    1   CAPT RON SMALL                    Value =  21.85   % =   .26
```

ID Number =	4	CAPT JERRY GERACE	Value =	19.21	% =	.23
ID Number =	10	MR. SWAMI KRISHNASWAMI	Value =	9.47	% =	.11
ID Number =	16	CAPT JIM HOLTEN	Value =	7.10	% =	.08
ID Number =	11	CAPT FRED STIERWALT	Value =	4.26	% =	.05
ID Number =	18	CAPT RIC ROUTH	Value =	4.25	% =	.05
ID Number =	7	CAPT DAVE HUNSUCK	Value =	3.73	% =	.04
ID Number =	3	CAPT MAX HALL	Value =	3.38	% =	.04
ID Number =	2	CAPT BOB RUSSEL	Value =	2.18	% =	.03
ID Number =	20	MRS. EDIE ROUTH	Value =	2.13	% =	.03
ID Number =	5	CAPT TOM GRIFFIN	Value =	2.01	% =	.02
ID Number =	17	DR. BILL CZELEN	Value =	1.84	% =	.02
ID Number =	8	CAPT MARK CLIFFORD	Value =	1.33	% =	.02
ID Number =	6	DR TERRY SKELTON	Value =	.50	% =	.01
ID Number =	12	CAPT MIKE HUNSUCKER	Value =	.49	% =	.01
ID Number =	14	CAPT DAVID KING	Value =	.33	% =	.00
ID Number =	13	CAPT PHIL FITZJARREL	Value =	.06	% =	.00
ID Number =	9	DR. WOODROW W. BLEDSOE	Value =	.03	% =	.00

Y standard deviation = 2.7

ID number = 1

Therefore, if we accessed the location 41,16 in window one's Recognition Database file, we would find:

X standard deviation = 1.3

Y standard deviation = 2.7

ID number = 1

(An example of the output from this process is shown in Table 5-2.) It is also possible to select or de-select records for training, allowing the user to do "what-if" testing. Once the coordinate database has been trained, it is ready to "recognize" an individual.

Identification. To identify a person, the "unidentified" person's picture must first be processed for gestalts, as previously described. At this point, the user selects "RECOGNIZE A PERSON" from the program "MAIN" on the ECLIPSE computer.

Using the gestalts, the program generates an ordered list of candidates, using the process described in chapter 4. The top person on the list is the winner (see Table 5-3.)

Since all processed pictures are stored in the "Processed Picture Database," the system can also load the

data for any of the pictures so that it appears to the system as if the picture had just been processed. This is done by selecting the option "LOAD A RECORD." The next section discusses the specific program modules used in implementing the system.

General Description of Program Modules.

The following is an overview of the different files and sub-programs used in the system.

FACE.MC -- The macro file on the NOVA which contains all the commands needed to run the face recognition system on the NOVA (see figure 5-11.)

RUNFACE.MC -- The macro file on the ECLIPSE which runs all necessary sub-programs on the ECLIPSE (see figures 5-12 through 5-14.)

GETFILE -- Requests a filename from the user for the facial image to be processed, checks that the file exists, and stores the filename for later use.

ADDFSTOP -- Converts older picture files into a newer format. Otherwise, it ignores the file.

TITLE -- Clears the monitor screen, displays a background image, and prints the filename at the top of the screen.

PROCESS1 -- Displays the requested file on the top left corner of the monitor, and stores the F-STOP value to a file on disk. Next, it performs an initial contrast-expansion of

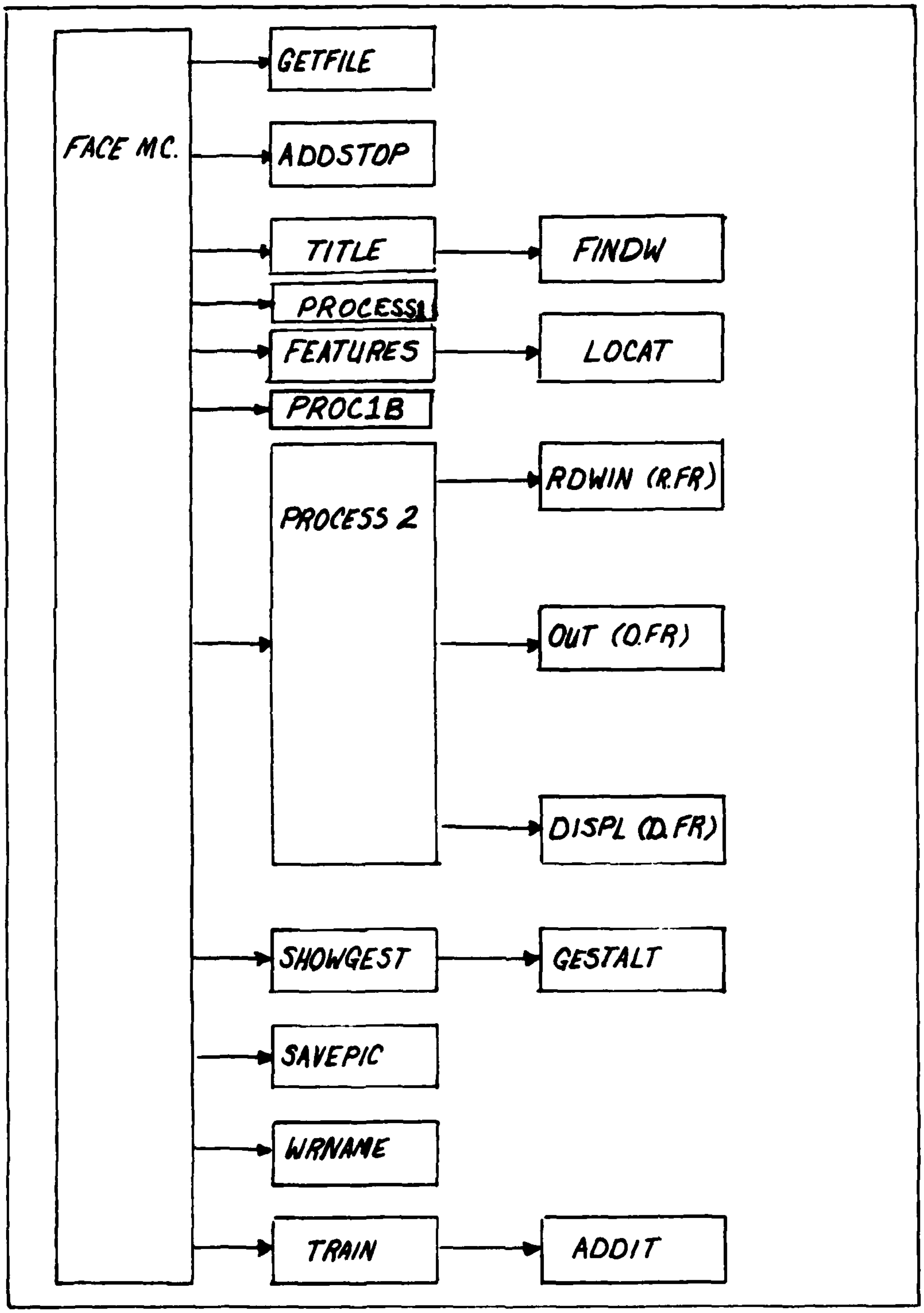


Figure 5-11. FACE.MC Program Calls (Runs on NOVA)

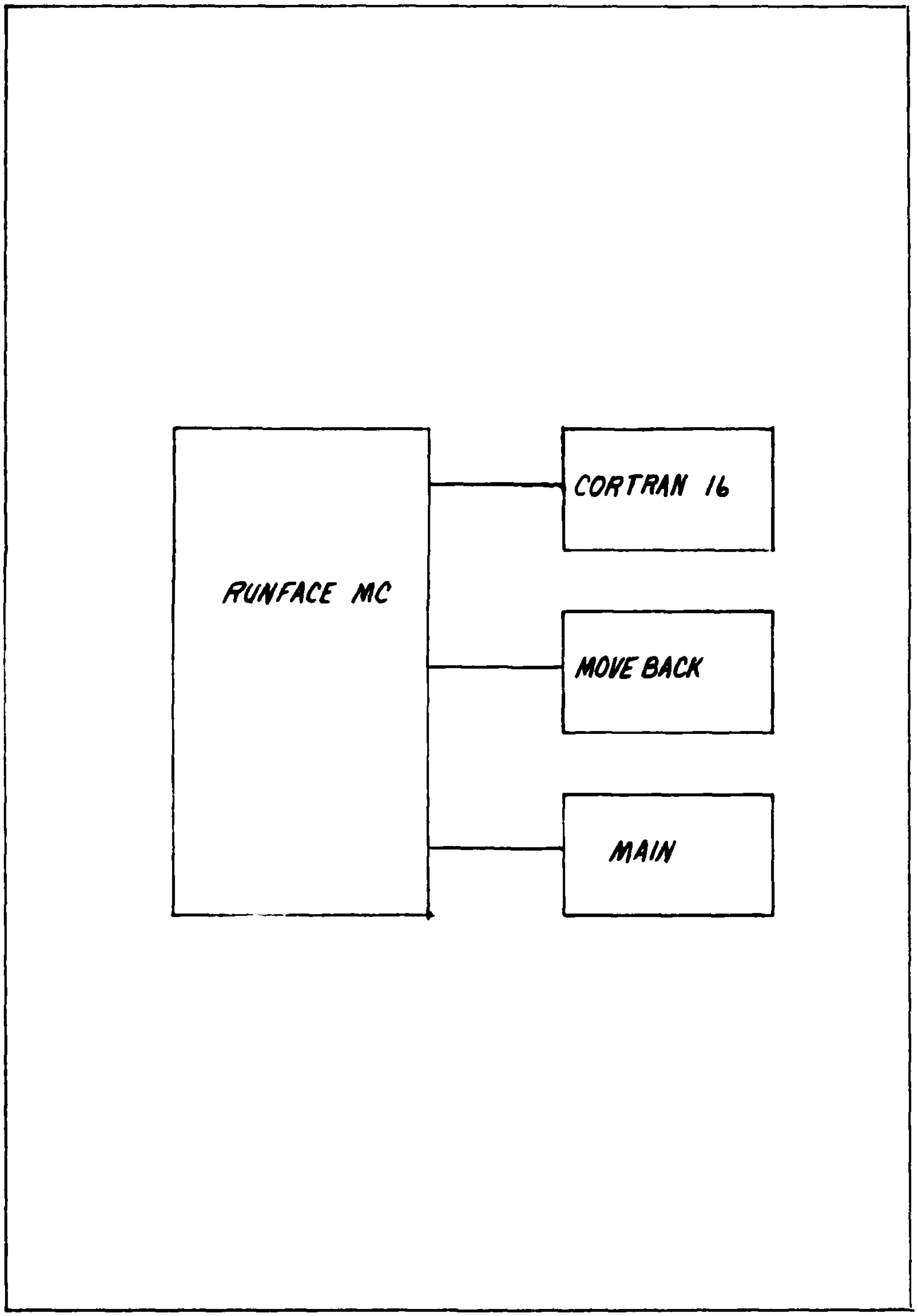


Figure 5-12. RUNFACE.MC Program Calls
(Runs on Eclipse)

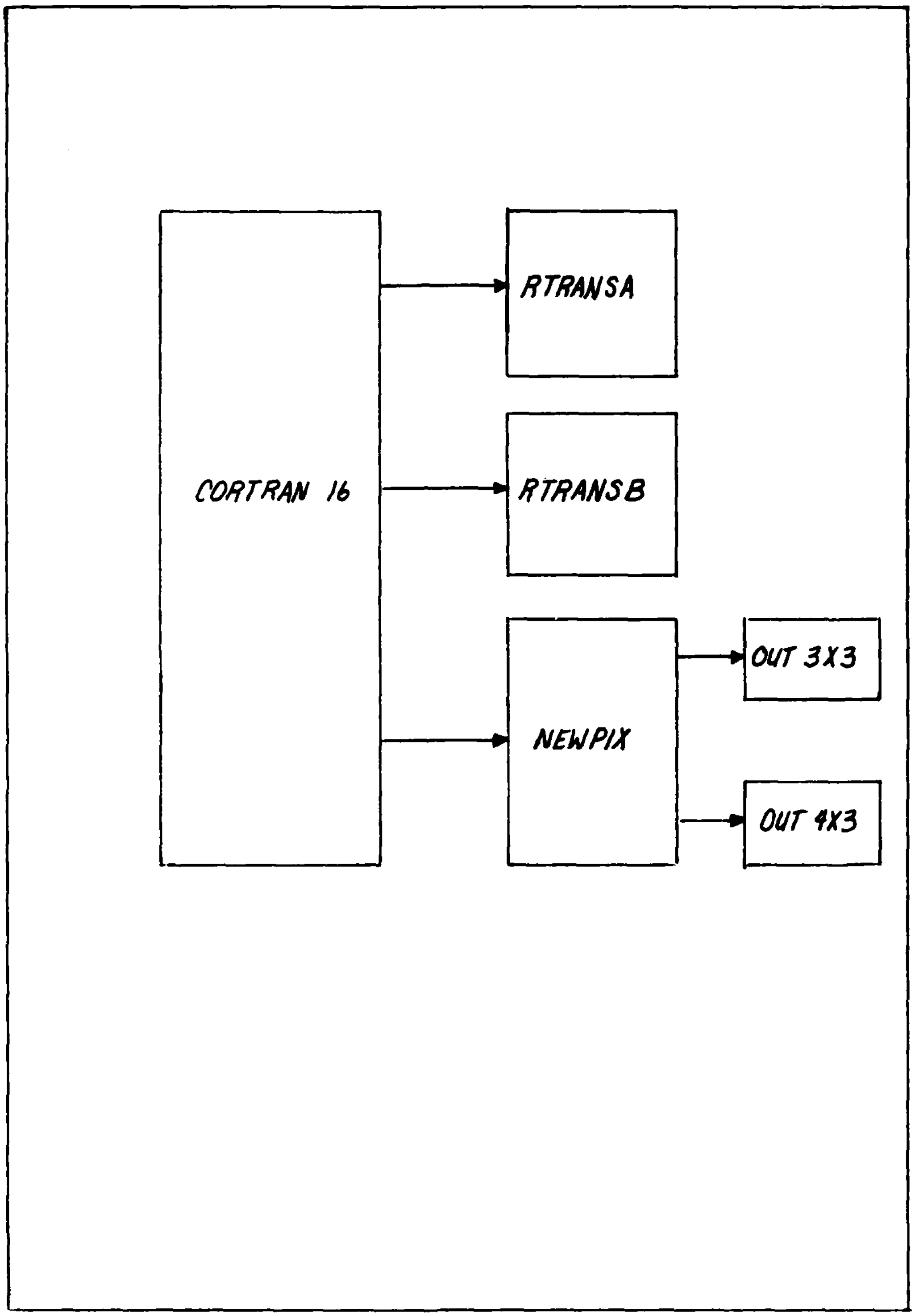


Figure 5-13. CORTRAN 16 Subroutine Calls

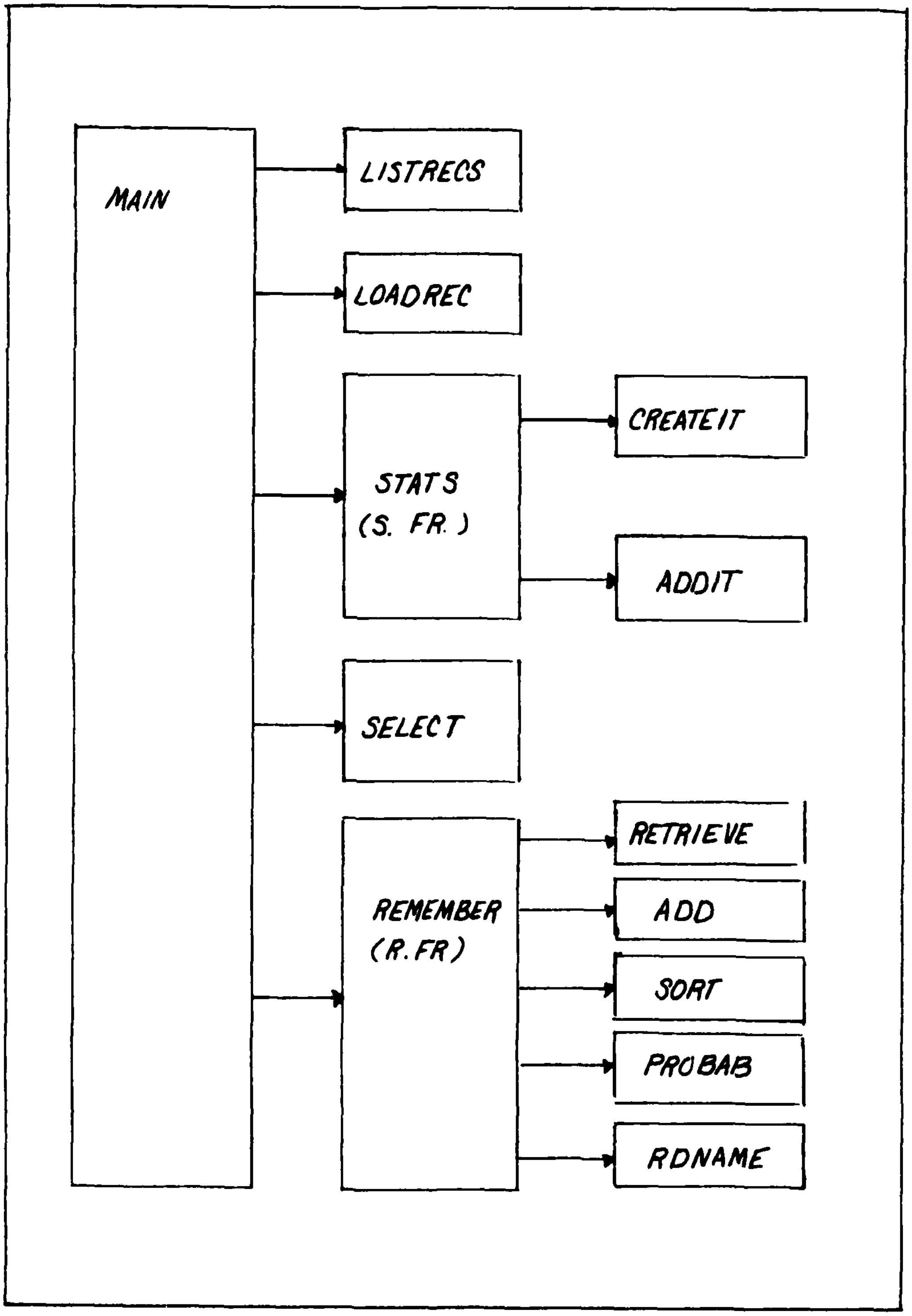


Figure 5-14. "MAIN" Subroutine Calls

the image, and displays the new image in the top-center of the monitor. Finally, it determines the X & Y locations of certain key features on the face (top of eyes, sides of head, etc), and displays an image at the top right corner of the monitor which has lines displaying the key locations. The locations are stored on disk in the file "WINDOWLOC".

FEATURES -- Allows user modification of feature locations, and re-displays locations on the screen.

PROC1B -- Performs final contrast enhancement of image, based on feature locations. Displays contrasted image. Stores contrast multiplier value to disk.

PROCESS2 -- Based on feature locations stored in the file "WINDOWLOC", extracts sub-images of the face from the final contrasted image. These images are stored on disk, and displayed on the monitor. (The images on the monitor vary from those on disk only in that the areas outside the partial face image are gray, not white, so as to emphasize the boundaries of the image.)

CORTRAN16 -- Calculates gestalts of the images sent to it from PROCESS2. Creates a file called "COORDPTS*.B, where the letter in the asterisk location denotes the number of the window file being processed (from 1 to 6.) For instance, the gestalt file for the sub-image from window 1 is COORDPTS1.B. CORTRAN16 runs in a loop, constantly searching for filenames from the NOVA. It terminates when the user is done processing pictures on the NOVA, and types

"QUIT".

SHOWGEST -- Displays gestalt values (which have been calculated by CORTRAN16) on the monitor, above the picture of the sub-image to which it belonged.

SAVEPIC -- At the user's option, will save the screen image to disk in a file called "TEMP.VD". The program will then create a file called "PRNTIMAGE", which signals CORTRAN16 to print the image in TEMP.VD. (This transfer of responsibility to CORTRAN16 was done because the print routine only works on the Eclipse.)

WRNAME -- Gives the user access to the "USER IDENTIFICATION" data for the following:

1) View all users and their ID numbers
2) Add a New User
3) Edit a User's Name

TRAIN -- Compiles a record containing all the gestalt values calculated for a facial image, and stores it in MAINPICS. Included are an ID number, filename of the original image, feature locations, contrast multiplier value, f-stop, etc.

QUIT -- Creates a file called "FACEDONE", which terminates CORTRAN16.

DETAILED DESCRIPTION OF SELECTED SYSTEM COMPONENTS

Much of the Face Recognition System will be self-explanatory to a reasonable programmer, but some areas need more explanation. This section will go into more

datail on the following:

1) Image File Format

2) Contrast Enhancement Process

3) Automatic Feature Location

4) Calculation of Gestalt Coordinates, and

5) Recognition Database

Image file format. The image within the box-cursor is stored in the upper left-hand corner of a 64x64 image file, with the rest of the picture filled with "white" pixel values (see figure 5-15). At the end of the file is then appended the X & Y window lengths, and the f-stop value of the picture (requested from the user by the program.) The window lengths are used later in the gestalt calculation for scale transformation, and then the f-stop value is stored with the gestalt values in a database. This format is used throughout this system for storing image files.

Contrast Enhancement Process. As discussed in Chapter 4, this system performs contrast enhancement on the entire face by the following process (see figure 5-16):

1) Measure average pixel value over a selected area of the image.

2) Multiply all pixel values in image until average pixel value in sampled area meets a specified value.

There are two contrast enhancement steps used in this system: one which samples a box located at a fixed location in the image (see figure 5-17), and one which samples a

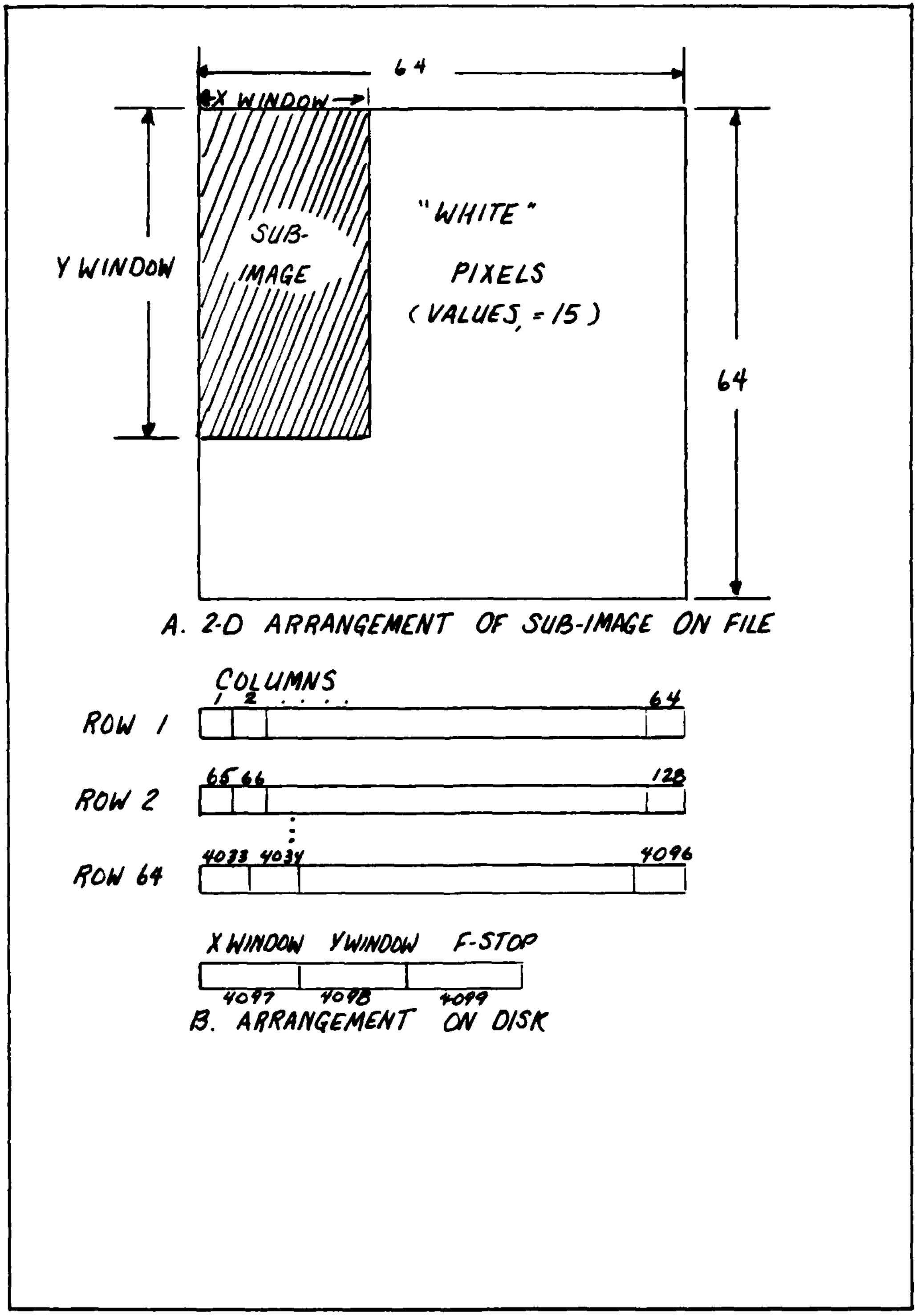


Figure 5-15. Image File Format

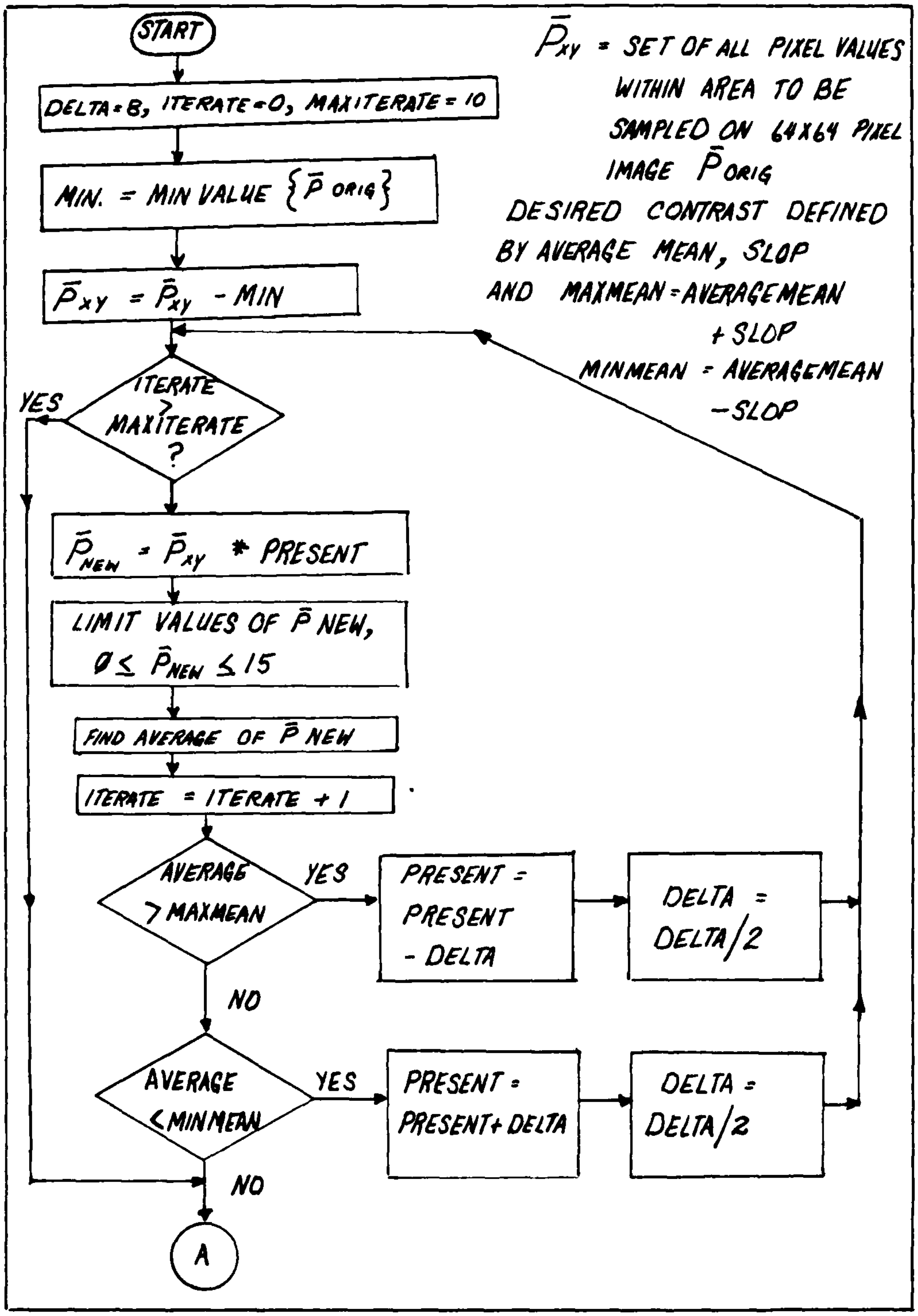


Figure 5-16a. Flowchart for Contrast Enhancement

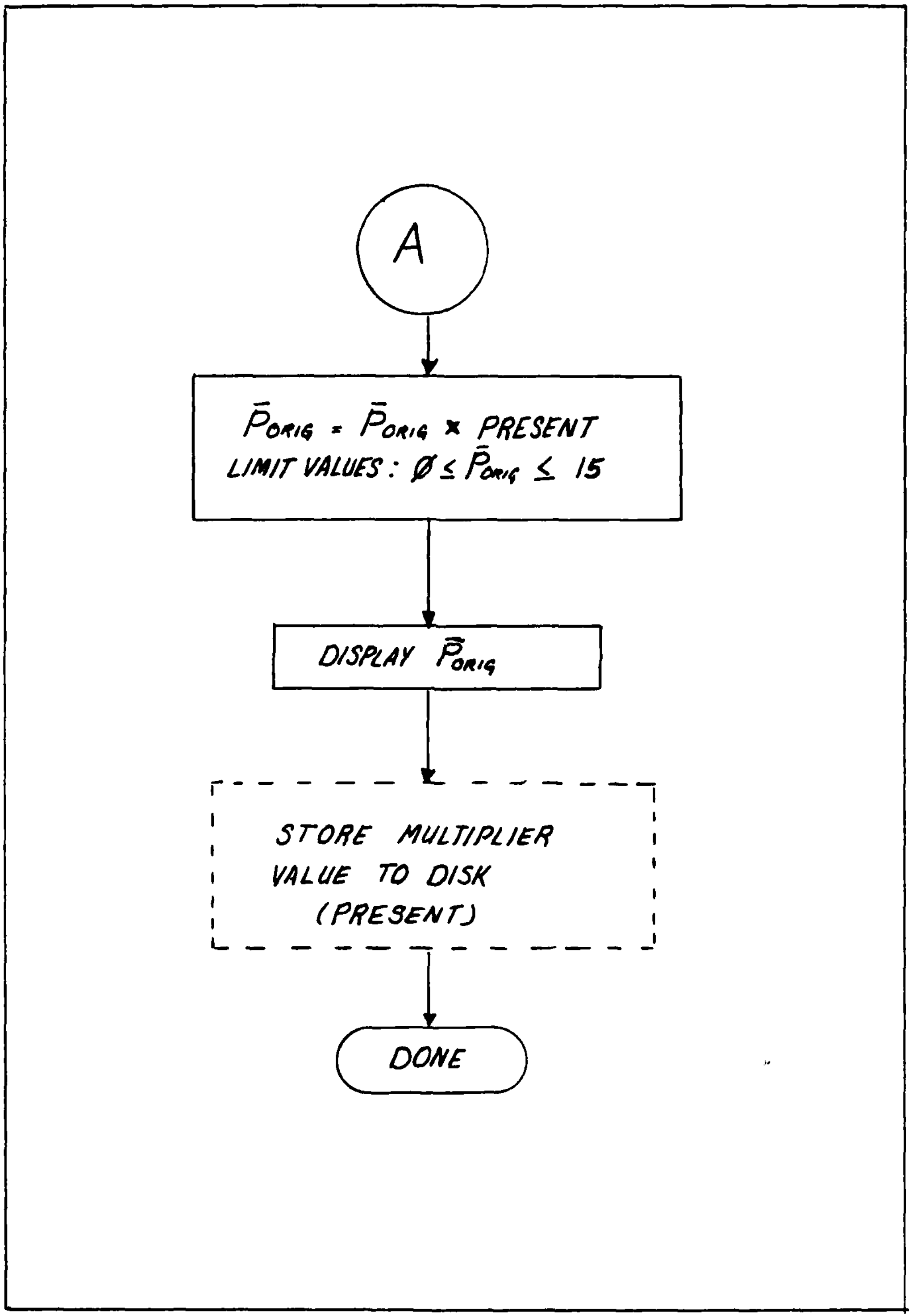


Figure 5-16b. Flowchart for Contrast Enhancement

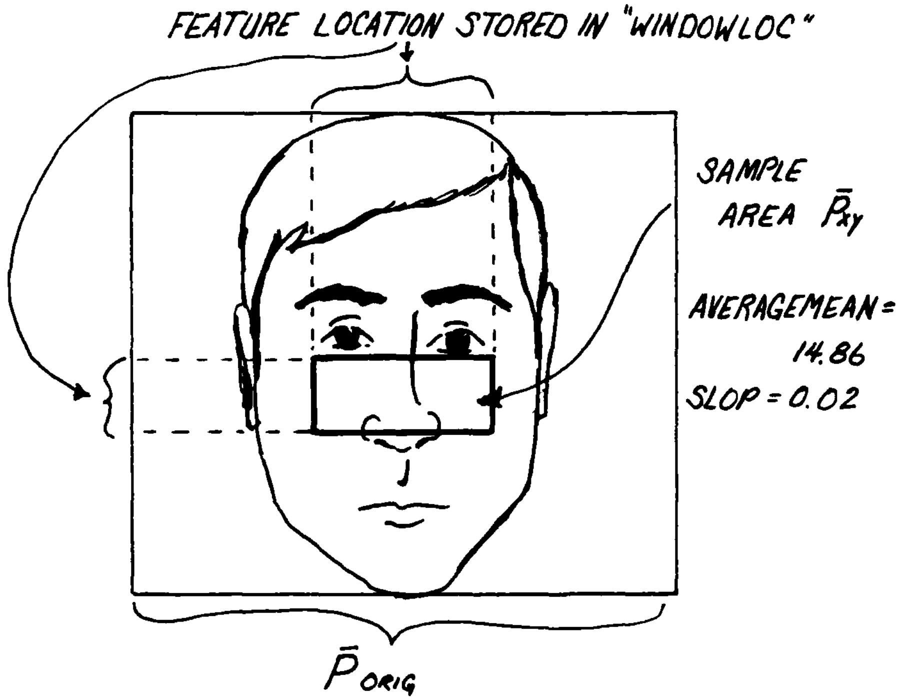


Figure 5-17. Sample Area and Constant for Initial Contrast Enhancement (in PROCESS 1)

rectangular area above the nose, and below and between the eyes (see figure 5-18.) Why use two steps instead of one? Unfortunately, the final contrast-enhancement process needs data from the feature-finder, but the feature-finder needs to work on a contrast-expanded image. The solution was to specify a fixed sample area on the image (even though its exact location on the face is not known by the computer), and perform an initial contrast-enhancement using this data.

Given the proper sample areas on the face, how can the proper contrast multiplier be found? Calculation of the contrast multiplier value is determined iteratively. Another way of saying this is to "try a value, and see what happens!" The system first tries adding a "delta" value of 8 to the initial multiplier value of 1. The system then checks to see whether the resulting average pixel value within the sampled range is above a specified range of allowable values, below the range, or within them. If it is within them, then the answer has been found. If the result is above the threshold, the system divides the delta value by 2, and subtracts this value from the multiplier. On the other hand, if the result is below the threshold, it also divides the delta value by 2, but instead adds this value to the multiplier. The process continues until either the average pixel value converges to the proper value, or the system has tried 10 times, at which point it is probably "close enough." (See figure 5-19.)

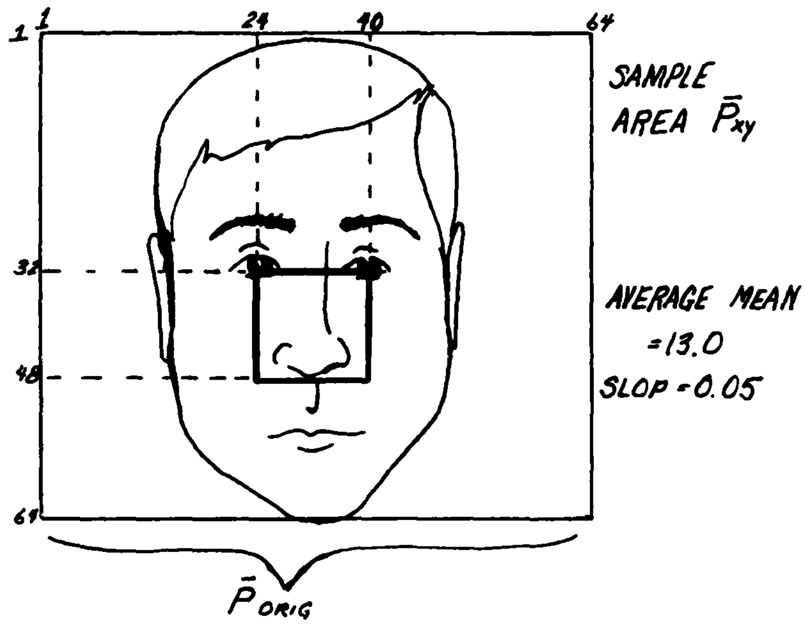


Figure 5-18. Sample Area and Constants for Final Contrast Enhancement (in PROC 1B)

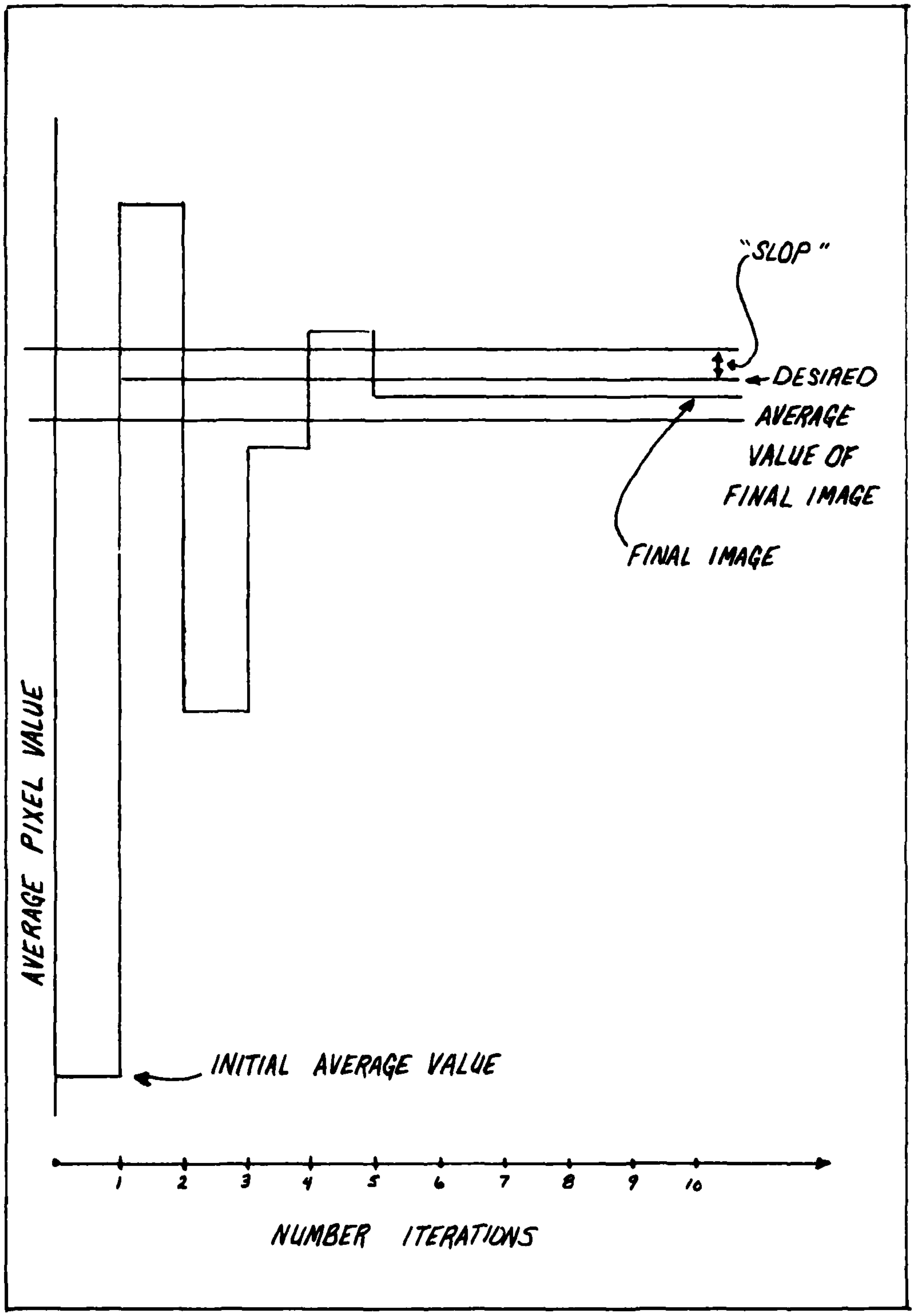


Figure 5-19. Example of Convergence to Desired Average Value

AUTOMATIC FEATURE LOCATION (Program "PROCESS1", with subroutine "FINDW").

The feature location process consists of four parts:

1) A knowledge base of allowable ranges within which different features may be found. (These ranges may overlap.)

2) A "boundary finder" which finds the boundaries of any high-contrast features within a given range.

3) A "box averager", which is used to estimate the center of a given area. An example of its use is in finding the center location between the eyes.

4) A set of rules used to determine the location of desired features based on resolution of the above data.

In order to find the location of features more accurately, the system first creates a high-contrast version of the image, as discussed previously. Once the system determines its estimates of locations, it creates a picture of the contrast-expanded image, with lines overlayed onto the image at the various feature locations, displaying it at the upper right corner of the monitor. In addition, it creates a disk file called "WINDOWLOC", containing the feature locations.

The feature finding subroutine FINDW initially stores more feature locations than are actually displayed and used by the system. (These other locations may not be accurate, as the user does not have the opportunity to update them

using the program "FEATURES", as he does the rest of the features.) The locations actually used are listed below in Table 5-4:

LOCATION IN FILE "WINDOWLOC"	FEATURE
1	Top of Head
3	Eyes Begin
4	Eyes End
5	Top of Nose
8	Center of Mouth
11	Chin
15	Center of Face (between eyes)
16	Left Side of Left Eye
17	Right Side of Right Eye
18	Left Side of Head
19	Right Side of Head

Table 5-4. Feature Locations used by System.

The horizontal locations (left side, right side, center of face) indicate the number of columns from the left side of the image. The vertical locations indicate the number of rows down from the top of the image. For example, if element 1 of "WINDOWLOC" has a value of 3, this means the top of the head begins on the 3rd row from the top of the image.

PROCESS1 changes a few of the feature locations found by FINDW before storing and displaying the values:

1) Top of Head

2) Bottom of Chin

3) Left Side of Head

4) Right Side of Head

The data for these locations is taken from the initial windowing the user performs on the original image. Many of these four features may become invisible after contrast expansion (particularly if the subject has light or gray hair.) Therefore, the system stores them prior to contrast expansion.

CALCULATION OF GESTALT COORDINATES

The foundation of the entire Face Recognition System is the calculation of the gestalt coordinate values. Appendix A discusses the form of this particular type of feature vector and why it is used, and chapter 4 discussed some modifications to the gestalt calculation. This section will discuss implementation of the gestalt process in this system by first showing how the one-dimensional gestalt transform is processed, then extending this process to a two-dimensional gestalt transform (from which the gestalt coordinate values are directly obtained.)

CALCULATION OF 1-D GESTALT TRANSFORM. The 1-D transform is calculated in two steps (see figure 5-20.) First, the coefficients for the necessary gaussian distribution are calculated by subroutine RTRANSA in the beginning of the program CORTRAN16. (In this way the coefficients need only be calculated once, regardless of the number of times used.) Subroutine RTRANSB then performs a series of correlations between an input array and parts of the gaussian

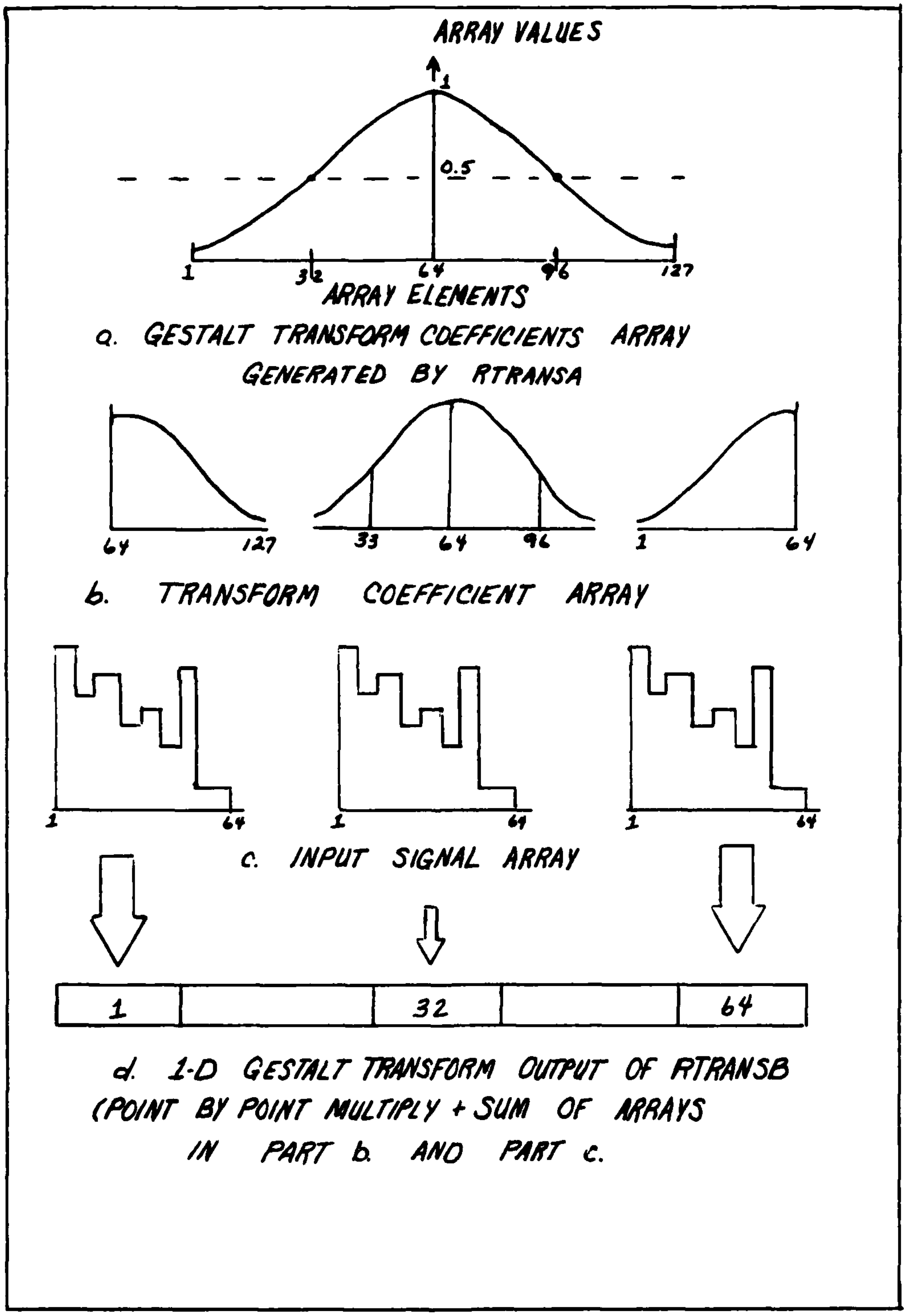


Figure 5-20. 1-D Gestalt Transformation Process Performed by RTRANSA & RTRANSB

distribution. RTRANSB will first perform a point-by-point multiply and add (dot-product) between the one-dimensional input array, and the right half of the gaussian distribution array (elements 64 to 127), putting the result in element 1 of the output array. To find output value 2, the range used on the gaussian distribution is shifted one to the left, so that elements 63-126 are being used. Once again the dot-product is taken of the input array and the distribution, and the result placed in element 2 of the output array. This process continues until the last element is calculated, using values 1-64 from the gaussian distribution. The output array is now available for use in the 2-D Gestalt Transform, to be described next. (This process provides an approximation to the particular 1-D Discrete Fourier Sine Transform used by Routh in his initial gestalt transformation experiments (1). The new transform also appears to be consistent with the physiology, as the structure of the cortex seems to imply that this is a process that would be trivial for the cortex to perform.)

CALCULATION OF 2-D GESTALT TRANSFORM. To find the 2-dimensional gestalt transform of an image, it is necessary to do the following (see figure 5-21);

1) Calculate the 1-D Gestalt Transform of each row of the image, substituting the result back into that row.

2) Calculate the 1-D Gestalt Transform of each column of the array resulting from step 1, and substitute the

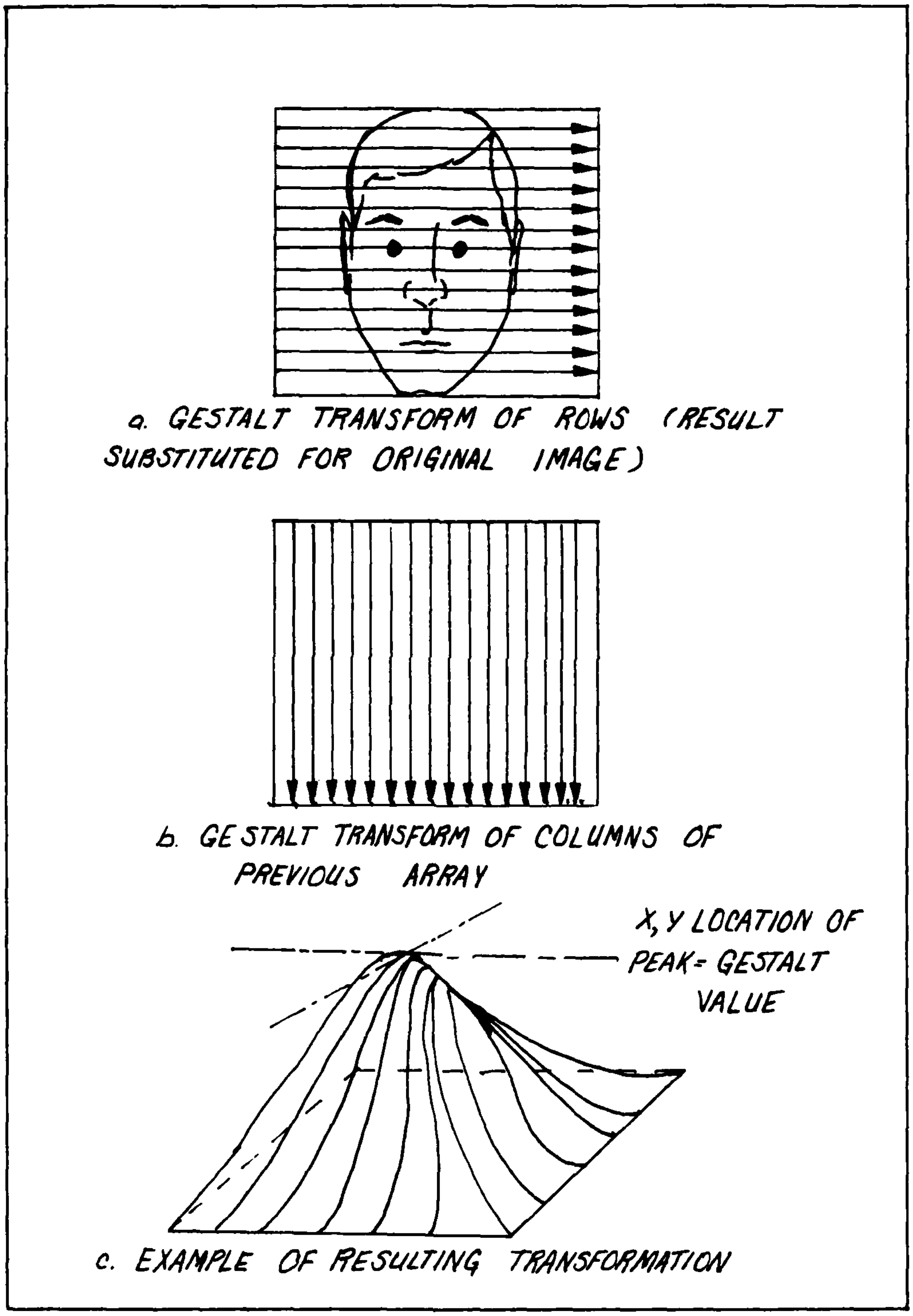


Figure 5-21. 2-D Gestalt Transformation Process

result back into that column.

(The concept of taking the transforms of the rows and then the columns for calculating a 2-D transform was taken from the same process used as a way of calculating a two-dimensional Discrete Fourier Transform (2D-DFT).)

The resulting 2-D Gestalt Transformed image will usually consist of a single hump (21). The X,Y array element containing the highest point on the hump denotes the values of the Gestalt Coordinates.

IMPLEMENTATION OF RECOGNITION DATABASE. (This implementation was co-designed with Dr. James R. Holten III.) The Recognition Database is a database containing the information on individuals for which the system is to be trained. It operates functionally like a 2-dimensional array of "stacks", where each stack can have any number of entries. Each entry is a record describing a single individual, who has been mapped to that location in the array. The database size is limited only by the amount of disk space in its directory on the computer.

There is a separate database for each of the six facial sub-image windows. Each database consists of three parts: the "Coordinate file," the "Lookup table," and the "Next free location pointer." The coordinate files are the six files "WINDOW1" to "WINDOW6". The lookup tables are the six files "WINDOW1.LU" to "WINDOW6.LU". The "Next free location pointer" files are the six files "WINDOW1.SP" to

"WINDOW6.SP".

The system "trains" the database by first determining the (X,Y) average gestalt values for the each individual, along with the (X,Y) standard deviation of gestalt values (as discussed under "Statistics Calculations" in chapter 4 and earlier in this chapter.) The (X,Y) average gestalt values determine the location of the individual's information in the database.

At this (X,Y) location is stored a record containing the ID number and the (X,Y) standard deviation, along with a field indicating the location of the next record at that coordinate value (see figure 5-22.) Any number of records can be stored at a particular coordinate value. The system can then retrieve all of these records just by specifying this coordinate value. The STATS statistics program, before it trains the system, completely re-creates the Recognition Database (as opposed to selectively updating an existing database.) It then adds a single record for each individual, as the Gestalt statistics for that individual are processed. Each individual only appears once in each window database. When trying to recognize an individual, the program retrieves records within a range of coordinate values, and then tests the records of just these retrieved individuals for recognition (as discussed in chapter 4.) Once again, each individual will only have one record which can be retrieved for any particular window database. The

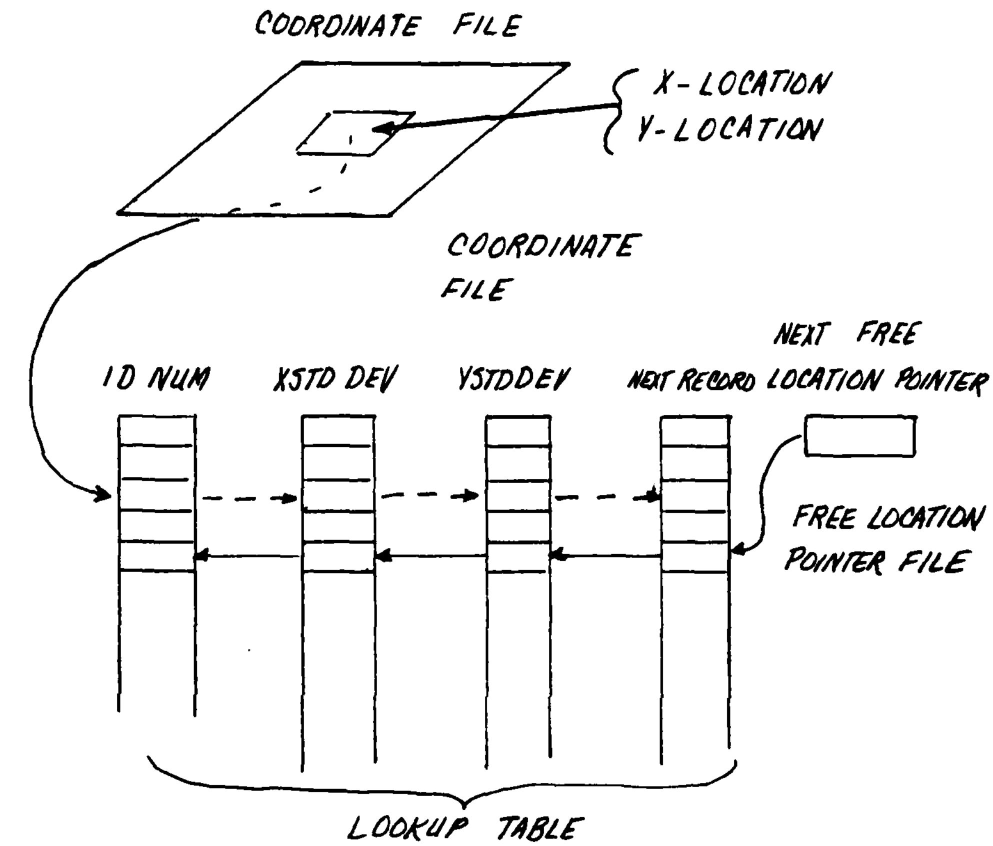


Figure 5-22. Basic Data Structure for Recognition Database

closeness of these retrieved values to the values of the unknown individual determines the machine-selected identity of the person.

There are two basic operations for this database -- adding and retrieving records. The implementation of these operations are discussed in the next two sections.

Adding a Record. Records are added to the system in the following manner (see figure 5-23):

1) The (X,Y) coordinate locations, obtained from the Gestalt Calculations, point to an entry value in the coordinate file.

2) The coordinate file entry is a pointer that points to a record in the lookup table. This record is the header of a linked list of lookup table entries stored at this (X,Y) coordinate file location. Each entry in the list is a record for a specific individual, including ID number, X Standard Deviation, Y Standard Deviation, and the pointer to the next linked list record, or zero for the last element.

3) The "Next Free Location Pointer" points to the next free record location in the lookup table.

4) The new record is placed at the location pointed to by the "Next Free Location Pointer."

5) The value in the (X,Y) location in the coordinate file is placed in the "Next Record" field of the new record, making the new record the head of the linked

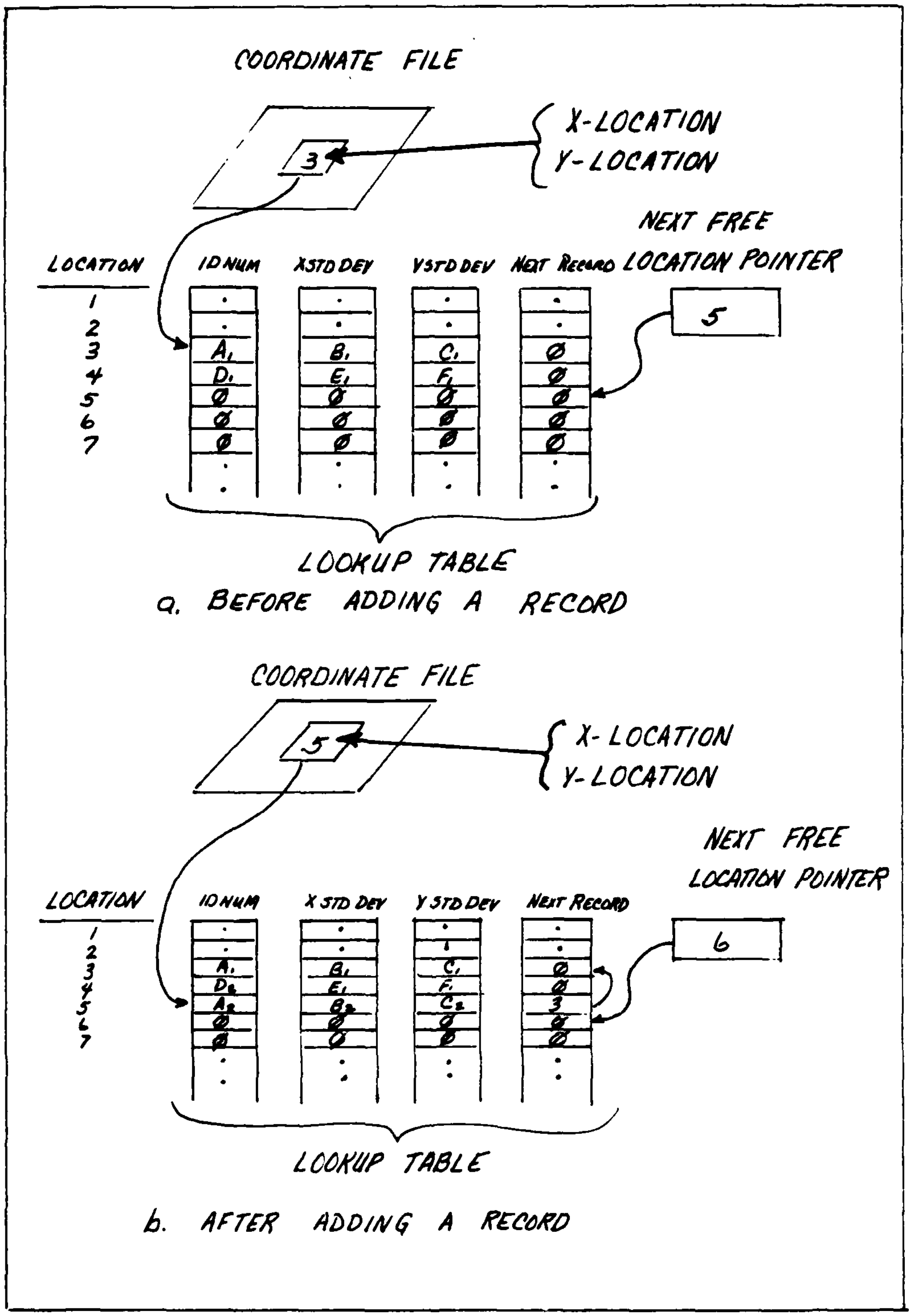


Figure 5-23. Adding a Record

list of individuals mapped to point (X,Y).

6) The location of the new record is placed in the (X,Y) location in the coordinate file.

7) The "Next Free Location Pointer" is incremented.

Retrieving a Record. Records are retrieved from the system in the following manner (see figure 5-24):

1) The (X,Y) coordinate locations specified by the program point to an entry in the coordinate file.

2) The coordinate file entry is a pointer that points to a record in the lookup table. This record is the header of a linked list of lookup table entries at this (X,Y) coordinate file location. Each entry in the list is a record for a specific individual, *including* ID number, X Standard Deviation, Y Standard Deviation, and the pointer to the next linked list record, or zero for the last element.

3) The record pointed to by the coordinate file entry is retrieved. If there is a number other than zero in the "Next Record" field of this record, then this number points to the next record to retrieve. When a record is finally encountered which has a zero in the "Next Record" field, then there are no more records to retrieve for this (X,Y) coordinate location.

RUN TIMES. The following are the average run times for key portions of the CTT Face Recognition System:

1) Processing a Picture for Gestalts: 8 minutes

2) Training the Recognition Database: 5 minutes

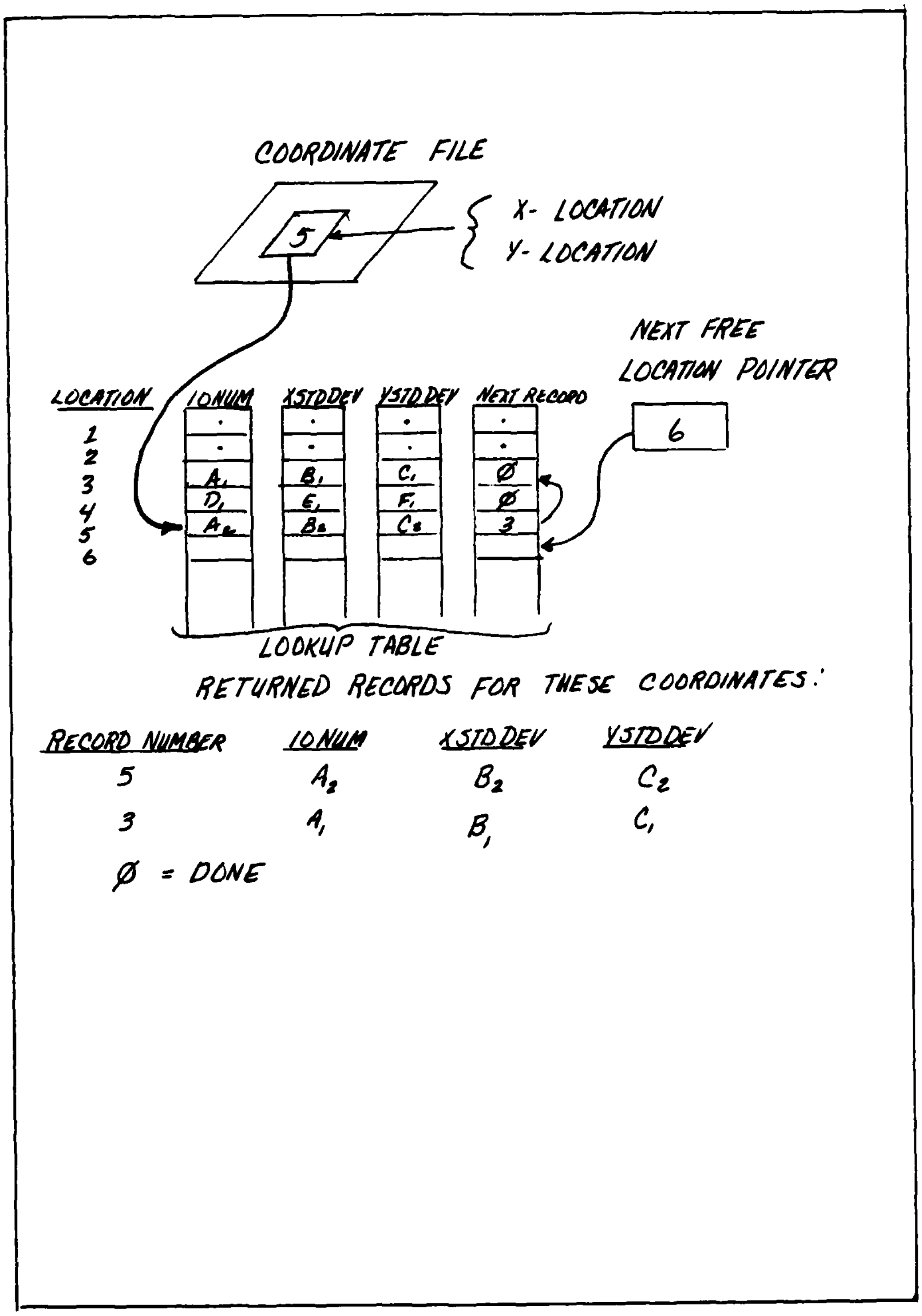


Figure 5-24. Record Retrieval

3) Recognizing a Person: 1 minute

These run times are for generally unoptimized code. Optimization would probably cut quite a bit of time from the process. In addition, as the entire system is inherently parallel, implementation on a parallel architecture, with dedicated hardware for gestalt calculations, would probably result in real-time processing for the most part.

SUMMARY

This chapter has discussed issues involved in actual implementation of the design which was specified in Chapter 4. The processing steps were discussed one by one, in the order they occur. This should give the reader the "big picture" of the operation. For those interested in more information, the actual program modules were discussed, with a detailed view of image file format, contrast enhancement, feature location, calculation of gestalts, and implementation of the RECOGNITION DATABASE.

VI. Testing, Results, and System Limitations

TESTING

The system was trained with from 4 to 9 pictures each of 20 individuals. For each picture, six sub-images were extracted, and the gestalt coordinate points calculated. Plots of these points are shown in figures 6-1 through 6-6.

The X & Y mean and standard deviation was then calculated for each set of prototypes, for each of the six sub-image windows, and the Recognition Database was built with this data.

One image for each individual was used to test the system. (This picture was not included in the training set.)

RESULTS

This section will discuss criteria for evaluating results, the actual test results, observations of performance in several areas, and a discussion of system limitations.

EVALUATION CRITERIA. The results were evaluated in two ways:

1) Percent absolute correctness

2) Average Reduction in Uncertainty

Percent Absolute Correctness is the percent that the individual being recognized appeared as first choice in the

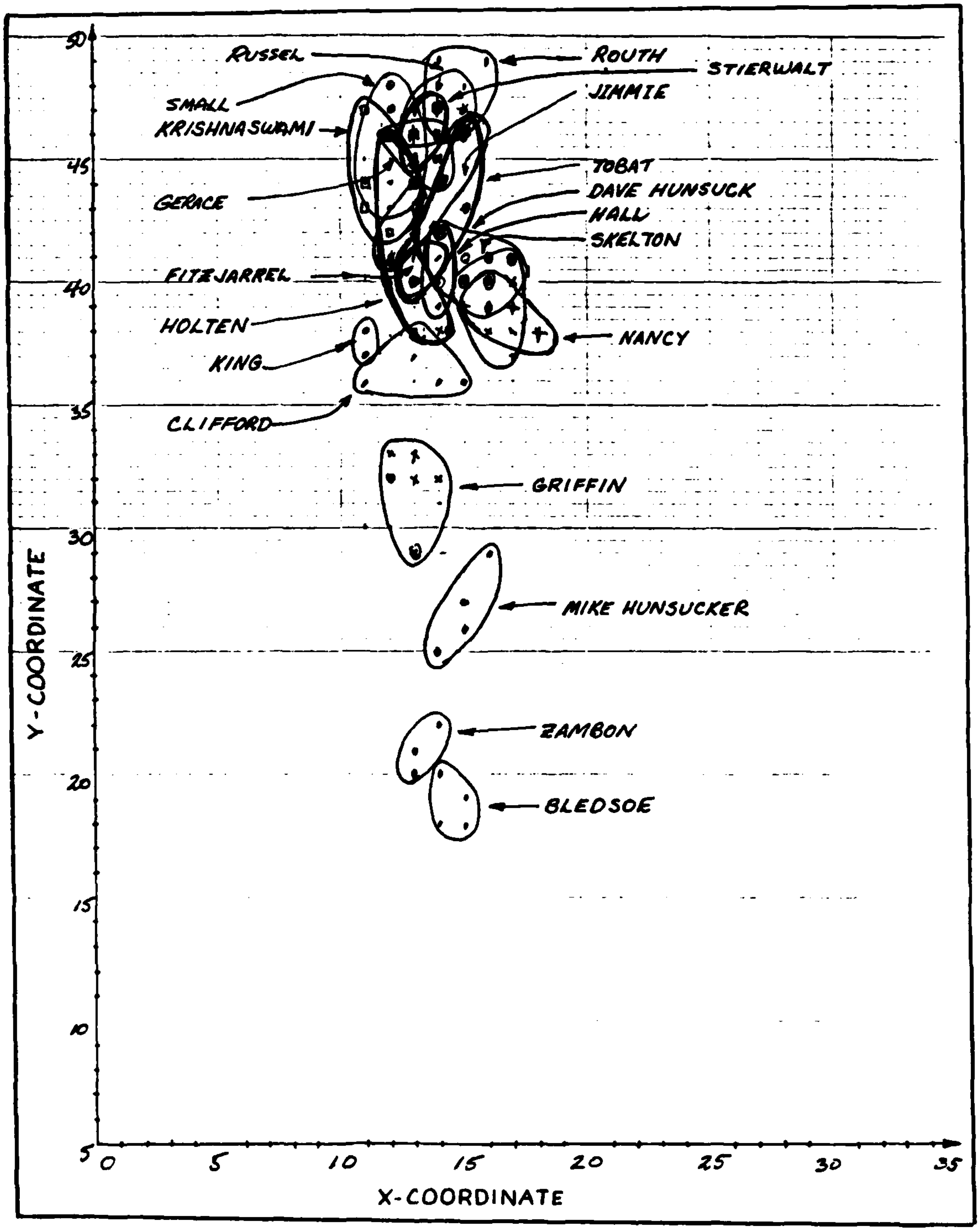


Figure 6-1. Window 1 - Left Half of Face

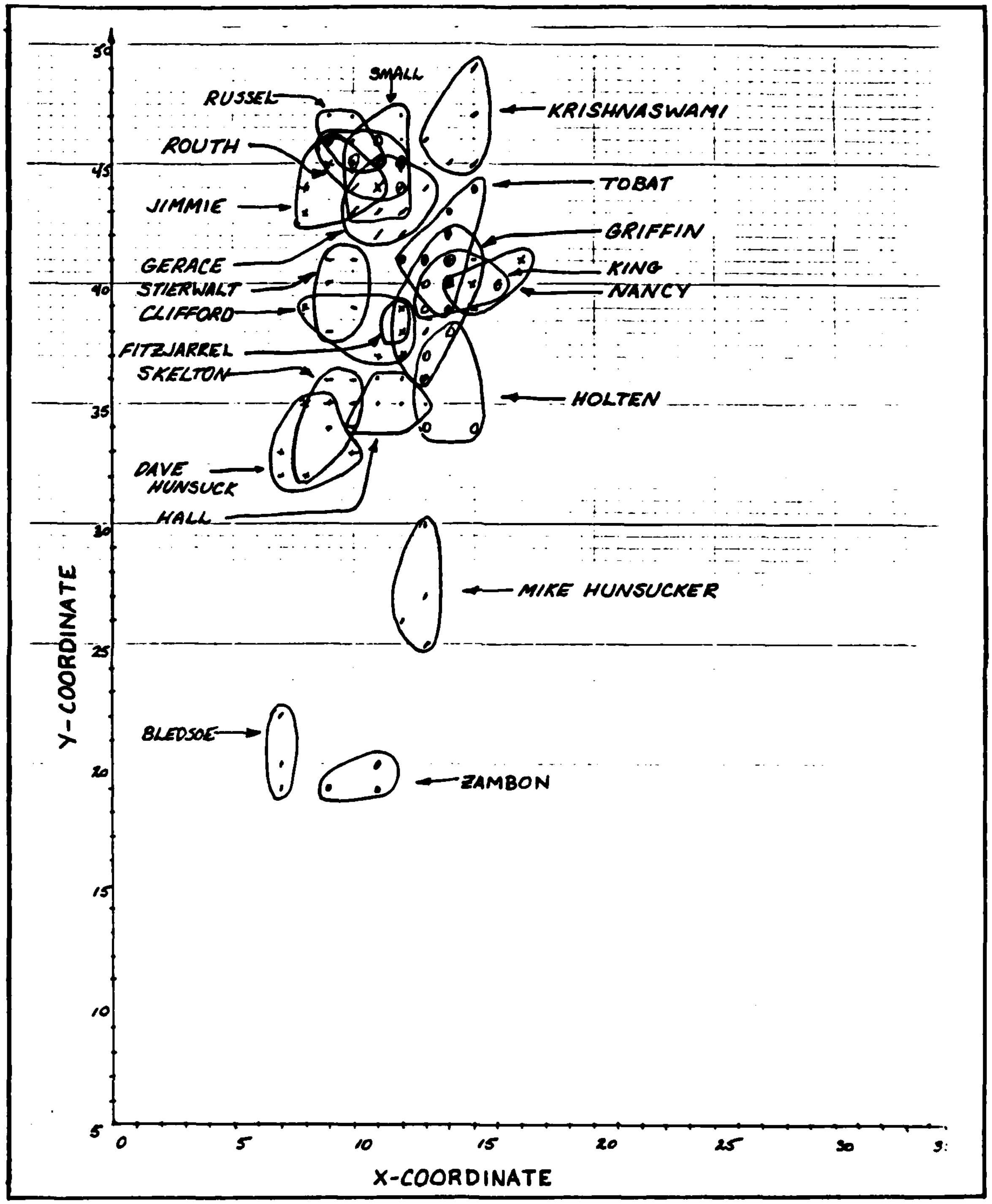


Figure 6-2. Window 2 - Right Half of Face

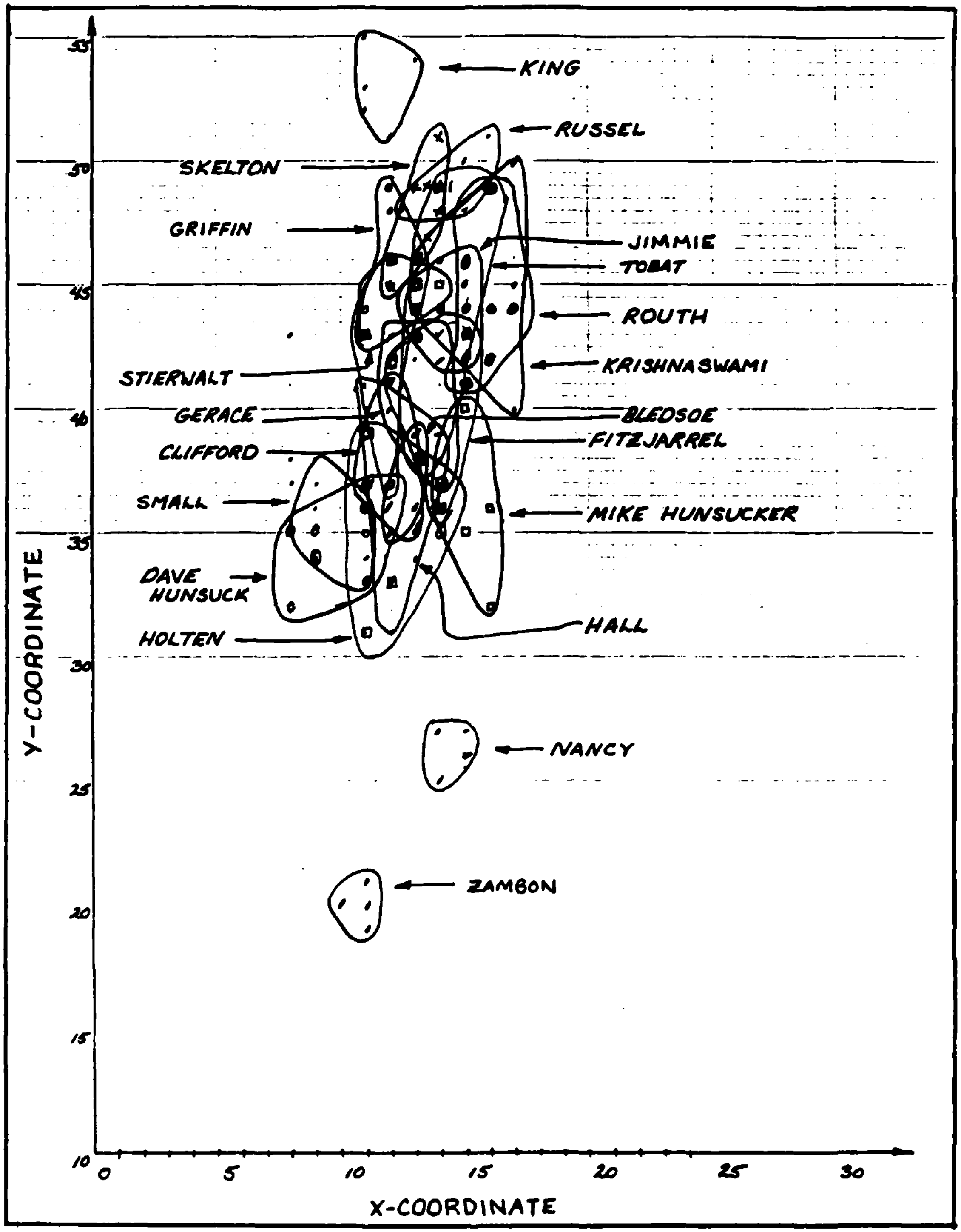


Figure 6-3. Window 3 - Top of Eyes to Bottom of Chin

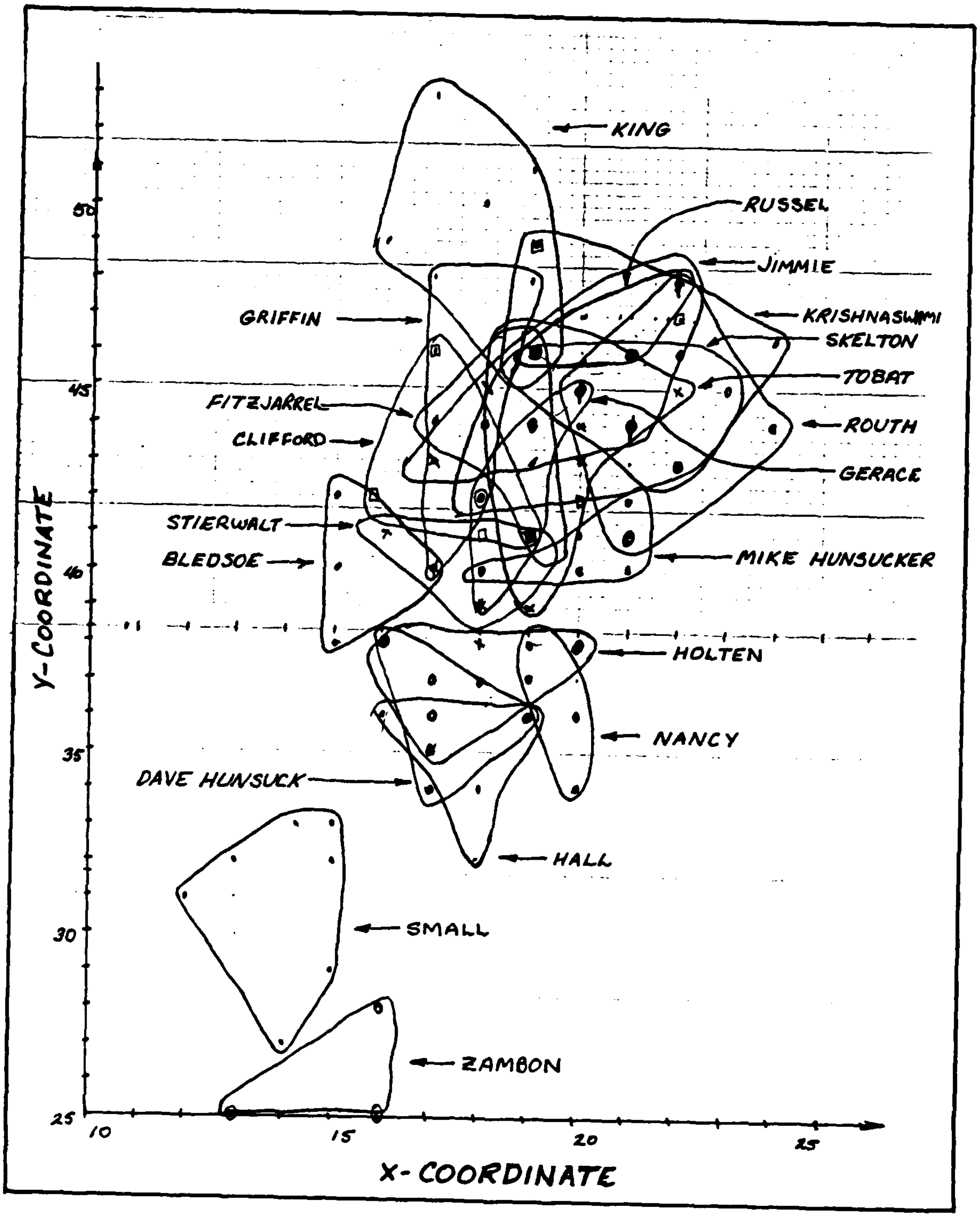


Figure 6-4. Window 4 - Top of Eyes to Center of Mouth

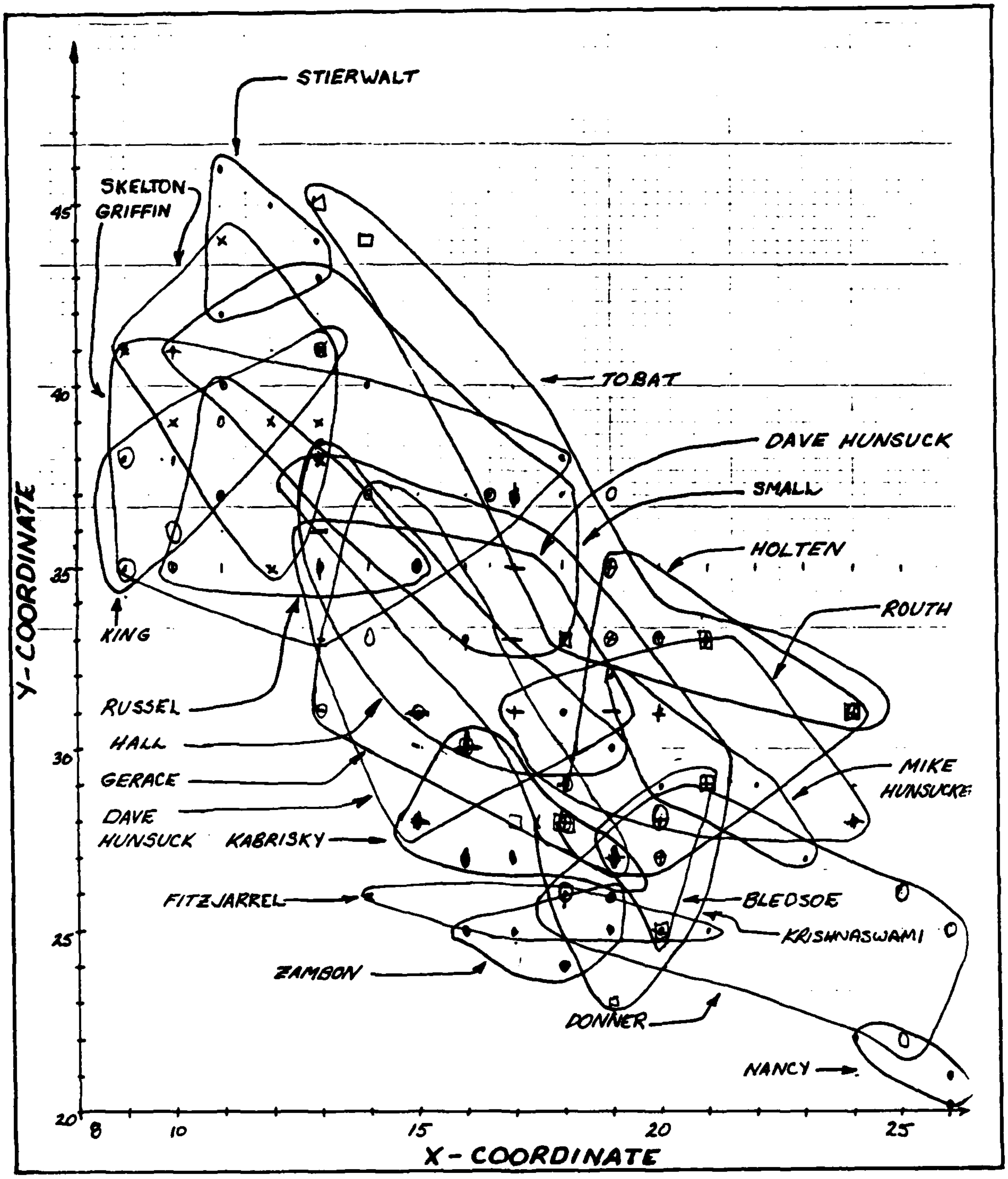


Figure 6-5. Window 5 - Top of Nose to Bottom of Chin

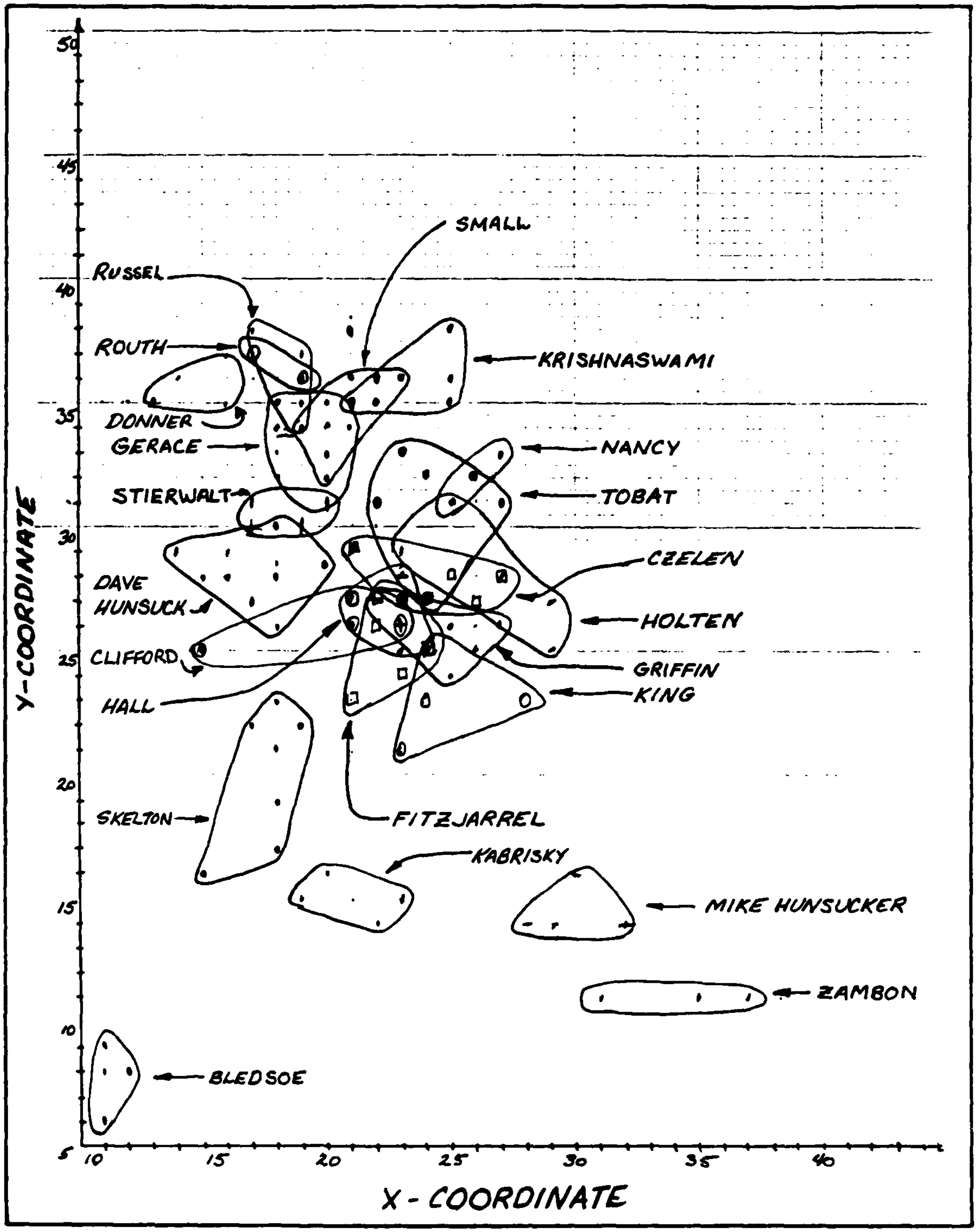


Figure 6-6. Window 6 - Top of Head to Bottom of Eyes

candidate list.

Average Reduction in Uncertainty is a technique used by Bledsoe in evaluating his results (2). It indicates how close the correct person was from the top, even though they may not have been chosen as first choice. It was adapted for use in this work as follows:

$$F = 1 - (1/M) \sum_{i=1}^{M} (Si-1)/N \qquad (6\text{-}1)$$

where S_i = number of individuals the correct person is down from the top of an ordered list of candidates

i = number of the particular individual in the database who is being processed for recognition

N = total number of individuals in database

and M = number of individuals for which the recognition system was tested

For example, if the correct individual is 6th out of 25 in the recognition list, the Reduction in Uncertainty is

$$R = 1 - (6-1)/25 = 0.80$$

If the correct individual is 1st, the Reduction in Uncertainty is

$$R = 1 - (1-1)/25 = 1.0$$

This measurement technique is useful, because it indicates the increase in information gained, even though the #1 choice may not be correct. The Average Reduction in Uncertainty is the result obtained when averaging the reductions in uncertainty for a number of individuals.

TEST RESULTS. The overall recognition results obtained were as shown in table 6-1:

Number in database:	20
Number recognized as 1st choice:	18
Number recognized as 2nd choice:	1
Number recognized as 3rd choice:	1
Absolute Correctness = 0.90	
Average Reduction in Uncertainty = .9925	

Table 6-1. Test Results for Recognition

1) Performance of the Individual Windows. This performance is shown in table 6-2:

Individual Window	Absolute Correct	Average Reduction in Uncertainty
1	0.50	0.915
2	0.75	0.983
3	0.60	0.870
4	0.35	0.933
5	0.30	0.823
6	0.55	0.958
All Combined	0.90	0.993

Table 6-2. Test Results for Individual Windows

The results from table 6-2 are plotted in figure 6-7. The data indicates that, although the individual windows had relatively low performance as far as absolutely correctness, the correct answer was usually close to the top, as indicated by the Average Reduction in Uncertainty.

2) Performance of Multiple Windows. In order to find the effect of incrementally adding additional windows to system, the recognition data was recalculated as the number of windows was increased from 1 to 6, going from the

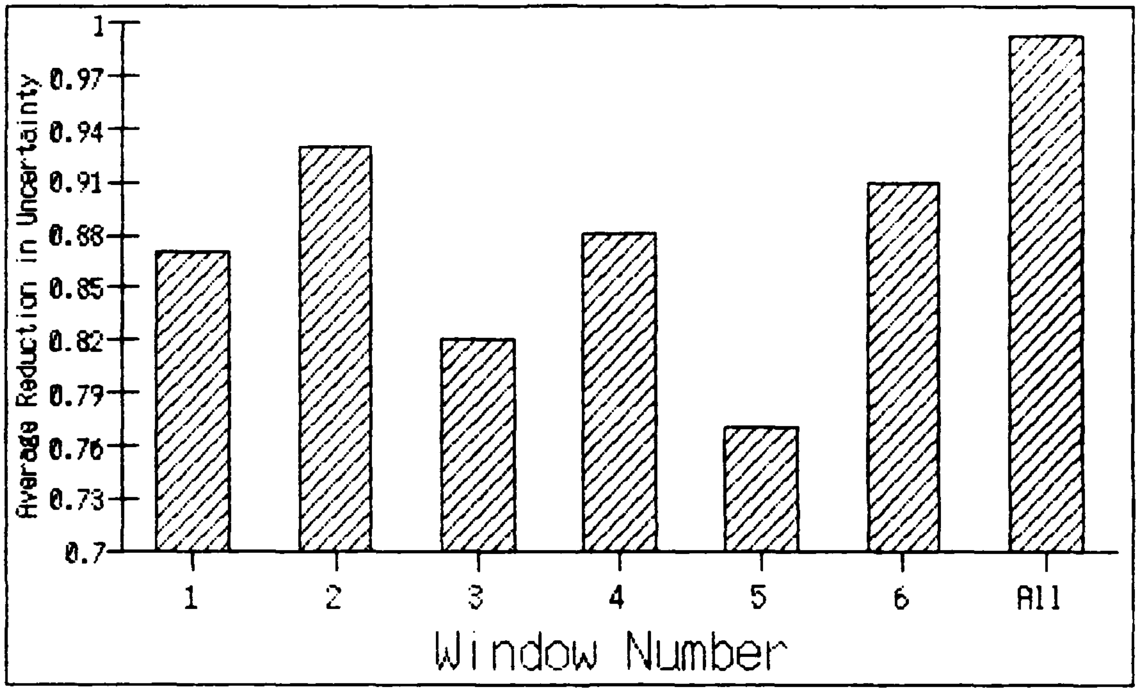


Figure 6-7. Plot of Actual Recognition Performance (by window)

best performing window to the worst. (The rank ordering of the windows was that which was calculated by the computer, as opposed to that calculated by actual recognition performance.)

Windows Used	Absolutely Correct	Average Reduction in Uncertainty
1	0.50	0.910
1,6	0.65	0.930
1,6,2	0.70	0.980
1,6,2,3	0.85	0.9875
1,6,2,3,4	0.90	0.9925
1,6,2,3,4,5	0.90	0.9925

Table 6-3. Recognition Results from Combining Multiple Windows

These results are plotted in figure 6-8.

There are those who would question the validity of Cortical Thought Theory on the basis that the CTT gestalt operation, which results in only a 2-dimensional vector, could not provide adequate resolution for a high-quality form recognition system such as is found in the human eye-brain system. This research, however, highly suggests that the gestalt operation as proposed by CTT can indeed provide high-performance form recognition, when coupled with the use of multiple windows on an image.

COMPARISON TO HUMAN PSYCHOLOGICAL RESULTS. The test results of this study were compared to those obtained from

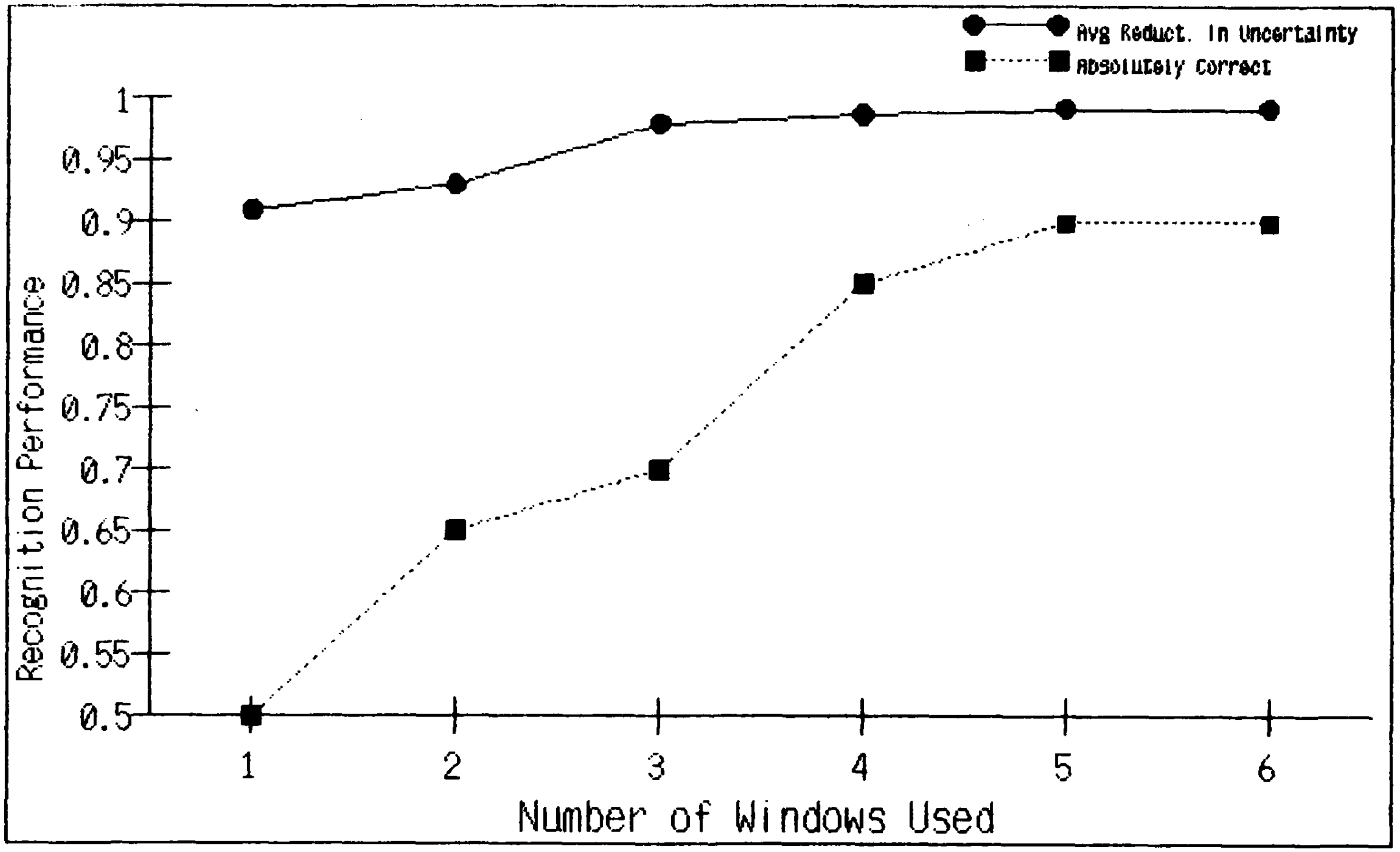


Figure 6-8. Recognition Performance as More Windows are Added

human psychological studies (as described in chapter 2.) This section particularly emphasizes a comparison, between the computer and humans, of which windows on the face provide the most recognition information. If the windows that perform high on this system also perform high in humans, and the windows that perform low in this system also perform low in humans, then it could be suggested that the feature vector set used in this system provides a valid model for human recognition performance. The CTT system results and the human results were indeed found to be quite similar, as described below:

1) Recognition Studies of Partial Faces. Chapter 2 describes an experiment where children were tested for their recognition performance when shown different portions, or "windows", of the face (8). Only half the windows used in the human study were used in the CTT Face Recognition system, but some useful comparisons can still be made. For instance:

a) In the human study, the half-face presentation provided the highest recognition performance over any other partial face image provided. This was also true in the CTT system.

b) In the human study, the 2nd highest performance was found using the window from the top of the head to the bottom of the eyes. This was also true in the

CTT system.

c) The nose-mouth window provided relatively poor recognition performance in both the computer and the humans.

d) Human performance tests on windows for single features (eyes only, nose only, etc.) resulted in very poor performance. The same results were obtained in limited testing in the CTT system, which is why none of these windows were included in the final set.

2) Recognition Performance for Babies. As described in chapter 2, studies with babies have indicated that the hair and the eyes held the babies' attention the most (9). The author makes a reasonable assumption here that the babies' attention to this part of the head indicates that this "window" provides the babies with the most information for recognition of the face. The CTT Face Recognition results also indicate a high amount of information for this part of the head. Another result of the baby experiment was that the babies decreased their attention to the mouth and increased the amount to the eyes and head when the mother was talking. In the CTT system, it was found that the already poor recognition performance using the window for the mouth area decreased with variation in mouth position or changes in expression. On the other hand, the window not including the mouth varied little with

a change in expression, as would be expected. Therefore, as found in the baby study, the CTT system would indicate a shift of importance farther away from the mouth area and farther toward the top of the head and the eyes.

3) Recognition with Expression Variations. A study by Galper indicated that human recognition performance was worse when the photograph of a person who had one expression on his face was used as the training set, and a picture of the same person with a different expression was supposed to be found (4). The CTT system also had the same problem. A person could be trained into the database with all pictures having the same expression, and then come back a week later and the person might not be recognized properly. The system just had too tight of contraints on what it expected the person to look like. The slight changes in expression and hair made a slight variation in gestalt coordinates which was, unfortunately, outside the system's constraints for that person. This led to having the person who was being trained into the system make slight expression changes and hair changes for different training pictures. This gave the system a better range of realistic values for the person.

DIFFERENCE IN INFORMATION IN LEFT AND RIGHT HALVES OF THE FACE. This study frequently found a significant difference in the vertical gestalt coordinates for the left and right halves of the face. This was particularly true

for individuals parting their hair on the left or right side of the head, as opposed to the center. The system did not get as high a variation in the X coordinates as expected for all the windows, although window four had a range of over 20 points, as opposed to the other windows which usually only had a range of about 10 points. Overall, however, the results of this research support the contention that the split-image portrayal of a face increases recognition performance.

OTHER RESULTS NOTED DURING TESTING.

1) The system will identify a face with only partially-recognized facial images. In many cases, an individual did not even appear as a candidate in one or two windows, but was still identified as a result of strong performance in the other windows. The system was determined to provide a reasonable engineering approximation to the Goldschlager set completion process.

2) Some individuals were "more recognizable" than others, as most of their gestalt-coordinates were in less-crowded areas of the gestalt-coordinate space. By the same token, some individuals were easily confused with others, as most of their gestalt-coordinates were in crowded areas.

3) Gestalt calculations of negative images gave little separation, as the system paid more attention to the

skin than the hair or features. This is because the system only works where black-colored pixels have the high-energy content. Humans also seem to have problems recognizing negative images. If humans are edging or cartooning the image, as some researchers suggest, then a negative image would give the same result as a positive one. Since a human is indeed sensitive to negative images, the CTT model may provide a possible explanation of why this is so.

DISCUSSION OF SYSTEM LIMITATIONS

The following are limitations or problems encountered in design and implementation of this system. These are not necessarily considered to be detrimental, but rather provide more insight into the "boundary conditions" of the process. In addition, this study does not consider a limitation bad if humans also experience the same limitation, as the purpose of Cortical Thought Theory is to provide a model of the human cortical processing mechanism. In fact, the existence of similar limitations between man and the CTT model adds more support to Cortical Thought Theory.

1) Window Location Dependence. Form recognition using the CTT gestalt process is very dependent on being able to define reproducable "windows" on the image, from one image to another. Routh hypothesized a method by which the retina, lateral geniculate bodies, and the cortex might be handling the general windowing process, but it is probably

not directly implementable in computer hardware in the foreseeable future. The CTT Face Recognition System required the development of a reproducable windowing process for the specific domain of human faces. For other applications, domain-dependent windowing may also provide temporary solutions until a general windowing system is developed. The most profitable research area in this direction would appear to be computer vision systems which map the boundaries of regions in an image.

2) Sensitivity to Expression. Changes in expression could change the results quite a bit, as would be expected. This was particularly noticable with opening and closing of the mouth. This problem was partially corrected in this implementation by training the system with pictures having many slight variations in expression and hair style. In addition, having multiple windows tended to reduce the effect of an expression change, because some windows were not as affected as much as others. If, for example, all of the subjects were to vary their mouth position widely during training, the system would calculate a large standard deviation for the windows sensing the mouth. It would then decrease their performance factors with respect to the windows which were relatively independent of the mouth. (In this system, window #6 (top of head to bottom of eyes) was most insensitive to the mouth, followed by window #4 (top of

eyes to center of mouth.)) Thus windows containing the mouth would be given less importance. A more advanced system would need many windows looking at multiple parts of the face to achieve increased independence of expression.

Is this an unreasonable limitation? The author thinks not, as there is evidence that in recognition studies that humans are also sensitive to changes in expression of people they are trying to recognize, if they have only seen that person with one expression (8).

2) Sensitivity to Scale. Although the gestalt calculation was designed to be scale invariant, the author has noted that this was not always the case. The gestalt calculation used in this system is less precise for a small image which is then scaled to a large size, than for a gestalt calculation on a full-sized image. For example, if an image is 1/4 full size, then the gestalt calculation which results from the smaller image would be multiplied by 4 to estimate the value for the full-sized image. As can be seen, the potential error grows greater as the image becomes a smaller percentage of the full-sized image.

The problem was minimized in this system by always making the original image as large as possible. The camera zoom was used to adjust the image size of the person, until the top and bottom of the head fit exactly within the top and bottom boundaries of the full-sized 64 by 64 pixel

window.

Much greater precision could be gained by taking pictures at a much larger size than needed, and then reducing the required partial face image down as required to fit it within a 64 by 64 pixel window. Therefore, there would be no loss in resolution due to scale for the gestalt calculation.

People with Dark-Rimmed Glasses. The most noticable problem with contrast expansion was with individuals wearing dark-rimmed glasses. The sample area used in this system for contrast-expansion happened to overlay the bottom of the glasses frame for individuals wearing glasses. For wire-rimmed glasses, the effect on contrast-enhancement appeared to be negligable. However, when the person had dark-rimmed glasses, the dark frame had a significant effect upon the average pixel value within the region, causing the contrast-expansion process to "over-expand" the image, washing it out.

In addition, it is unclear to the author how to properly window the eyes when the person has dark-rimmed glasses (i.e., window around the rims or the glasses or on the eyes themselves through the glasses?) At present, a subject who wears dark-rimmed glasses is asked to remove them prior to training or recognition.

Sensitivity to Rotation. A question which has

frequently arisen during this study is "What happens if you rotate the head? Can the system still recognize it?" The answer was found to be no, except for small rotations (about plus or minus 5 degrees.) The reason for this is that there is apparently significant new information displayed on the head for every 10 to 15 degree rotation. Unless the system has been trained for every 10 to 15 degree rotation, it has no a-priori knowledge of the information. (The requirement for full-face and side-face photos in mug shots are due to this problem. The author, however, would maintain that even the information in these two pictures are not sufficient for a human to recognize the person at any angle.)

The following is a possible way to implement a system which is reasonably independent of head rotation. Dr. Woodrow W. Bledsoe, during his experiments in face recognition in 1966, developed a system which could estimate the number of degrees rotation of a human head (Bledsoe, p 10.) Using such a system, the CTT Face Recognition System could use the number of degrees rotation as an index to the proper database for that range of rotation.

Other problems to be resolved would be how to properly define window locations and properly contrast-expand facial images with different rotations. The rest of the CTT system could be essentially the same as presently designed.

Is the present sensitivity to rotation a problem? Not

necessarily, as there are many applications where the user could restrain the rotation of the person to be recognized. For instance, the system could be used to compare the front view of a suspect with the front view of subjects in a mug-shot file. (Of course, adding a database for the profile view would greatly increase the performance.) Another application would be security access, where the subject could sit in front of a camera in order to be recognized.

The bottom line is that extension to a reasonable rotation invariance should be relatively straightforward, but the present system is more than adequate for many applications.

SUMMARY

The CTT Face Recognition System was performance tested with a database of 20 people. The following are some of the significant results:

1) It identified the correct person as 1st choice 90.0% of the time, and the Average Reduction in Uncertainty was 99.25%.

2) The six individual windows had relatively poor performance when taken individually, but when combined achieved the above-stated performance. As a result of the promising results from combining the windows, the combination mechanism developed in this study is suggested

to be a reasonable engineering approximation of the Goldschlager Set Completion Mechanism.

3) The recognition performance of the windowed images on the face were quite similar for both the CTT system and humans, suggesting that the feature vector set used in the CTT Face Recognition System provides a valid model for human recognition performance.

4) The performance of this system can be extended as needed by increasing the number of sub-images processed on the face. In addition, the operations of the system are inherently parallel, giving the capability of "real-time" processing with any number of windows.

VII. Summary and Conclusions

SUMMARY

A face recognition system was developed, based on the principles of Cortical Thought Theory (CTT), recently developed by Dr. Richard L. Routh at the Air Force Institute of Technology. CTT claims to be a generic model for sensory information analysis, regardless of the domain or entry level of abstraction. Routh tested the CTT architecture successfully for speech processing. In order to test this architecture as a generic model, CTT was tested for visual processing, specifically for the difficult task of human face recognition.

As an initial test, the 64x64 primary audio cortex map was removed from Routh's speech system, and in its place was inserted a 64x64, sixteen gray level, digitized image of a human face. This analysis was applied to five images each of sixteen different people. The results indicated that human faces can be classified and distinguished with the CTT model, and the 2-D CTT mapping (or "gestalt") of the faces is psychologically similar to the way a human would group them.

Work continued on an advanced face recognition system. In this system, pictures were contrast-enhanced automatically by the computer to increase recognition performance and allow use on people with different skin

colors. An algorithm was developed to find feature locations on the face. Using these locations, the system extracted six sub-images from the contrast-enhanced image, calculated the 2-D gestalt coordinates, and stored the information in a database. Statistics were then calculated on at least five prototypes processed for each person. Overall performance of different sub-windows on a face were also determined. "Unidentified" individuals were recognized by calculating the six gestalt feature vectors for their pictures, and then finding the closest match to previously stored data. The results of the individual windows were combined by using an engineering approximation to a "set completion" mechanism. This process identified the individual having the set of six feature vectors which most closely matched those of the unidentified person. The computer generated an ordered list of candidates by closeness of match.

A knowledge base was constructed of from 4 to 9 prototypes each of 20 different people. Performance testing of the system yielded a reliability of 90%. The performance of the individual windows was determined, giving insight as to what parts of the face provide the most recognition information. In addition, the cumulative effect of combining windows was shown to provide performance much greater than the individual windows themselves.

CONCLUSIONS

The system exhibits many characteristics of human recognition. The following are the significant results of this research:

1) Provides an explanation of why the primate visual system splits images vertically before displaying them on separate right and left primary visual cortexes.

2) Provides an explanation of why humans experience difficulty in recognizing negative images.

3) Maps faces which look similar to humans close together in CTT space, and maps faces which look quite different to humans far apart in CTT space.

4) Partial face images which seem to give the highest recognition performance in human psychological experiments give the highest performance in the CTT model.

5) The system is consistent with the human physiology as is presently understood.

6) Provides an engineering approximation to Goldschlager's set completion mechanism as interpreted by Routh.

7) Highly suggests that the gestalt operation, as proposed by CTT, can indeed provide high-performance form recognition when it is coupled with the use of multiple windows on an image. This is a result predicted by CTT and borne out in this research.

The performance of the face recognition system strongly

suggests CTT's general applicability to vision, and increases its credibility as a general model of human sensory information processing. The conclusion of this research is that Cortical Thought Theory is a promising new architecture with demonstrated effectiveness, worth increased research and development by those interested in developing computing systems with human-like sensory information processing capabilities.

VIII Recommendations

The following are recommendations for continued research and development in both face recognition and general visual image processing using CTT.

CTT Processing

1) Investigate windowing mechanisms for other form recognition domains, such as English letters, to gain more insight into development of a general windowing process.

2) Combine Routh's speech recognition system (which did not extract sub-images from the audio signal) with the visual windowing capabilities of the Face Recognition System to investigate correct window engineering on the audio spectrum plot for universal speech recognition.

3) Combine a CTT vision system with a vision system which locates the boundaries of objects in a scene, such as the system developed by Dr. James R. Holten III in his dissertation, "A Robot Vision System." His boundary-finding system, with further development, could possibly take care of much of the front-end windowing for the CTT system.

4) Investigate how gestalt values change with different gaussian distributions to determine the robustness of the gestalt transformation.

5) Integrate a high-level database language, such as dBase III or R-Base 6000, into the system to increase the

flexibility and query capability of the database.

Face Recognition

1) Improve the feature location algorithm to reduce the amount of operator intervention necessary. One suggestion would be to help estimate feature locations by using statistical data for feature locations determined in previously-processed pictures. (The operator corrects any discrepancies in feature location before a picture is processed, which means all of the previously-processed pictures will have the correct feature locations stored.)

2) Add more windows to increase the performance of the system. Recommend the following as a start (for both left and right halves of face)

a) Top of head to top of eyes.

b) Top of eyes to top of nose.

c) Bottom of eyes to bottom of chin.

d) Outside edge of eye to side of head

e) Bottom of hairline to bottom of chin

3) Investigate other methods of contrast-enhancement, which are not subject to the limitation caused by dark-rimmed glasses.

4) Determine the best windowing scheme for people with dark-rimmed glasses.

5) Increase the number of individuals trained in the database, in order to better quantify the recognition

performance of the system versus number of people in the database. In addition, quantify the proper number of windows necessary to identify a given number of people with a given accuracy.

6) Perform more human recognition studies using the specific windows used in the CTT Face Recognition System.

7) Investigate use of the CTT Face Recognition model to better understand and provide therapy for the disorder known as "prosopagnosia", or the inability to recognize human faces.

8) Estimate the maximum number of faces discriminable by the system as a function of its configuration.

Appendix A

Cortical Thought Theory: A Working Model of the Human Gestalt Mechanism. (22)

This paper is reprinted from a paper presented to the Aerospace Applications of Artificial Intelligence Conference, Dayton, Ohio, Sep. 85. It presents the overall concepts involved in CTT, particularly calculation of the "gestalt."

CORTICAL THOUGHT THEORY:
A NEW COMPUTING ARCHITECTURE
BASED ON THE HUMAN BRAIN

by

Richard LeRoy Routh, Ph.D.
Robert L. Russel, Jr. and
Matthew Kabrisky, Ph.D.

Department of Electrical Engineering
Air Force Institute of Technology
Wright-Patterson AFB, OH 45433
rrouth@wpafb-afita

I. INTRODUCTION AND SCOPE

This paper is a brief report highlighting some of the results of a fundamentally new approach to human brain modeling called Cortical Thought Theory (CTT). CTT is not a neuron based model of the human brain. The authors of this paper are skeptical of the approach taken by past and present attempts to discover the information processing architecture of the human brain by first exhaustively investigating neuron behavior, and then attempting to construct a computing architecture from these neurons. It is our opinion that such an approach is likely to be as unfruitful as an attempt by South American Bushmen to discover how a car works by first undertaking a thorough investigation of the electron shell properties of the metal in one of the pistons. Not only are they ill-equipped to investigate electron shell behavior, but the systems information about how the various metal components interact to propel the automobile is not contained in the complete knowledge of the electron shell behavior of the metal in the pistons. Likewise, it is our opinion that the systems information about how the human brain processes information is not contained in the complete knowledge of the function of neurons.

Instead, the approach used by CTT is the systems approach which attempts to show, through a top-down investigation of the system, the necessary form of the solution which specifies the information processing

computing-architecture used by the human brain. Perceptual psychological, neurophysiological and neuroanatomical data are used in this top-down systems analysis of human brain function, but they are used only as constraints which serve to narrow the theoretical form of the solution. By employing these constraints, and others from theory of computation, this new CTT approach shows the form of the solution is so narrow that we can make some useful statements as to the function of the cortex. Working from this base, further experimental investigation was suggested which resulted in the mathematical specification of the function of the cortex. A simulation was built which processed both audio (speech) and visual (human face) inputs. The resulting speech recognition machine performed in a manner which was psychologically similar to the human speech recognition system (HSRS). It also predicted a new class of audio-illusions which have subsequently been synthesized and verified as true human audio illusions. The resulting image recognition machine has a high reliability (91%) of distinguishing (identifying) any single human face from a data base of twenty different human faces. In addition, the CTT architecture accounts for multiple previously difficult-to-account-for human natural language phenomena, to include, among many important others, the ability of humans to apparently directly access the single most important inference or piece of information, regardless of the size of the knowledge base.

II. APPROACH

It appears that there are only two conceptually different mechanisms for reasoning. These are deduction and induction. Deduction is used to define a complete formal reasoning system which specifically prescribes the operators which may be used to relate the pieces of information in the knowledge base. A Turing-machine-like architecture is ideally suited for problems which lend themselves best to solutions using deductive mechanisms. Artificial intelligence offers many examples of limitedly successful attempts to model the human knowledge representation structure with deductive techniques. The problem with this approach has always been that search times increase exponentially with linear increases in the size of the knowledge-base.

CTT shows why the use of a Turing machine to model and access the human knowledge representation structure must necessarily result in this exponential explosion. Instead,

a reasoning mechanism which is inductive must be used. By induction, we mean a mechanism must be able to (1) analyze an input, (2) extract its gestalt (the essence of its form), (3) remember its gestalt, and (4) compare it to all other previously remembered gestalts to find and quantify the closeness of the best match in order to establish relevent associations. The general model of this process is shown diagrammatically as follows:

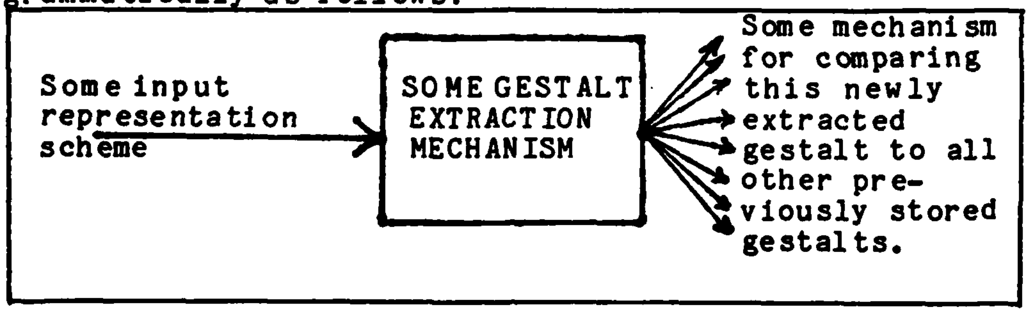


FIGURE 1.

By using constraints from several areas of science, a model can be developed for the cortex implementation of induction. This paper will concentrate on the input representation scheme, and the gestalt mechanism.

III. CORTEX MODEL FOR INDUCTION

Input Representation Scheme.

By using both neuroanatomical and neurophysiological data, it was possible to constrain input representation scheme of this general model for induction into the more specific input representation scheme for the induction mechanism used by the human brain. It was agreed that for all domains at all levels of abstraction in the human brain there exists a single standardized input representation scheme which presents any input as a two-dimensional image.

Gestalt Mechanism.

A knowledge of the theory of computation was used to show that the cardinality of the human gestalt feature vector set (GFVS) is two, regardless of the domain. It was also shown that a corollary of this argument provides an explanation of why <u>any</u> attempt to model the human knowledge representation and inferencing structure with a Turing machine must necessarily suffer from exponential explosion.

By using the experimental results obtained from the perceptual psychology investigations into the nature of the human gestalt mechanism by Kabrisky, Maher, Ginsburg, Pantle, and Sekuler (among others), it was argued that the two element gestalt vector is probably extracted from some low pass two-dimensional spatial frequency domain representation of the 2-D input image.

But what spatial frequency domain representation was to be used? Several methods of displaying the low-frequency spatial harmonics of a 2D-DFT were investigated so as to find a single identifying 2-space vector characteristic which could be called a "gestalt". The method had to suppress the D.C. value which did not contain useful information for identification.

It also had to deal with how to present both sine and cosine components of a 2D-DFT on a 2-dimensional surface. It was oberved that if the Two-Dimensional-Discrete Fourier Sine Transform (2D-DFST) was used (instead of the 2D-DFT), and if the technique of zero-filling was used to produce sub-integral harmonics, a "hump" was usually observed between the zeroeth and the first harmonic. The location of the peak of this hump could easily represent the gestalt value since it can be represented by a two-space vector, and it changes location for different input images (see figure 2.) Experiments suggested that it was sufficient to examine the 1/64th harmonics between zero and one. The 2D-DFST gestalt mechanism is specified by the following equations:

<u>Given the Discrete Input Image</u>: M_{kh}; $k,h=1,\ldots,64$,

<u>Then</u>: $$S_{kj}=\sum_{h=1}^{64} M_{kh}\sin\left(\frac{2\pi j(h-1)}{4096}\right) \quad ; \quad k,j=1,\ldots,64 \qquad (1)$$

$$T_{ij}=\sum_{k=1}^{64} S_{kj}\sin\left(\frac{2\pi i(k-1)}{4096}\right) \quad ; \quad i,j=1,\ldots,64 \qquad (2)$$

$$\Rightarrow \text{GFVS}=\left\{(i,j):\ \max_{ij}\ T_{ij}\right\},$$

where GFVS=(i,j) is the two-space vector identifying the location of the gestalt on the next higher (in the hierarchy of abstraction) local cortex surface.

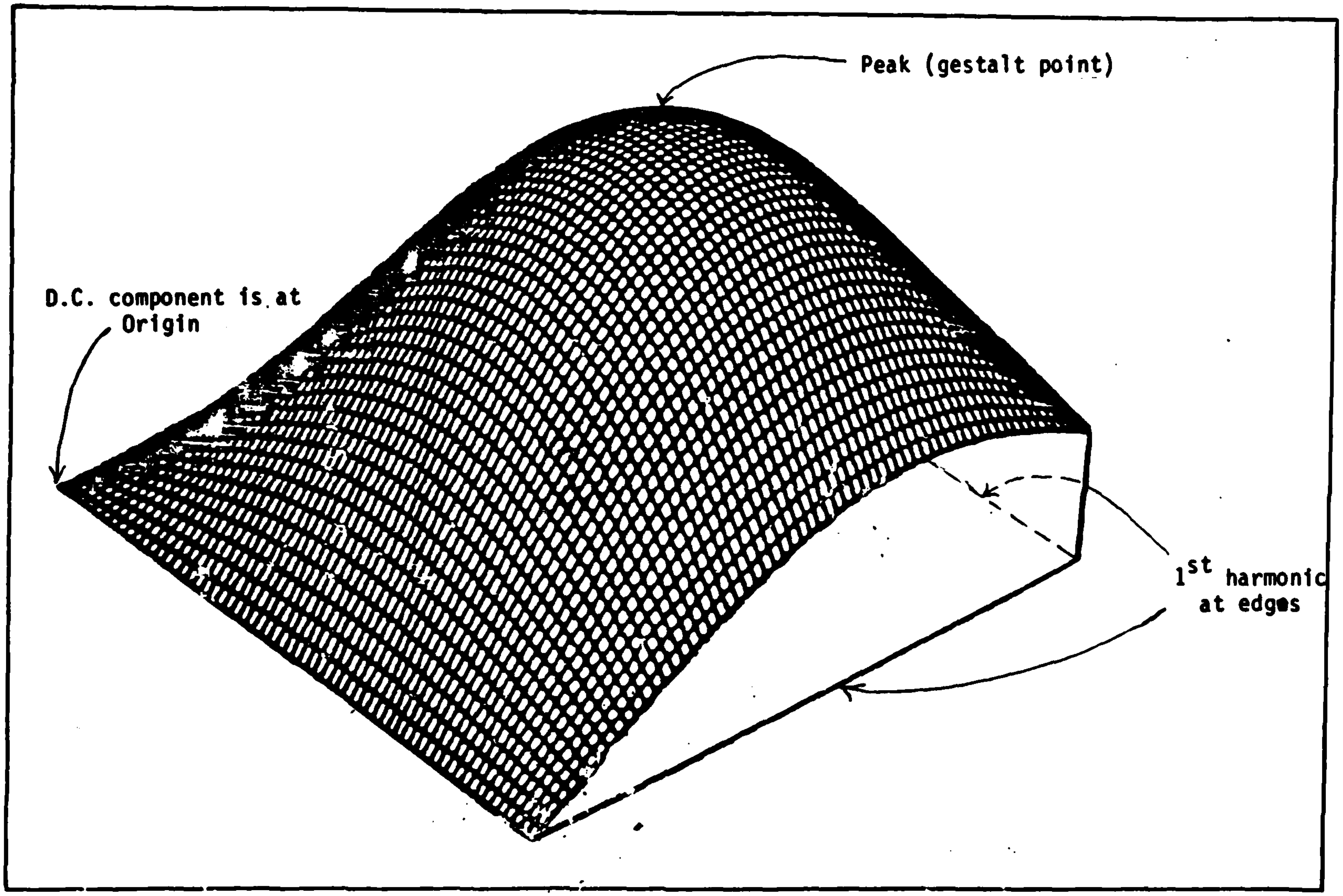

Peak (gestalt point)
D.C. component is at
Origin
1st harmonic
at edges

Figure 2.

It was shown that the level one neurons of the cortex could easily perform a very good approximation to the 2D-DFST from the zeroeth to the first harmonic. There would be an error between the true 2D-DFST and the cortex transform, but the cortex transform still preserves the important characteristic: it produces a "hump" whose peak moves in relation to the human-perceived difference in the input images. The gestalt would be two-space location of the cortical column located at the highest amplitude point.

The transform used to simulate this process is as follows:

<u>Given the Discrete Input Image</u>: M_{kh}; $k,h=1,\ldots,64$,

<u>Then</u>: $$S'_{kj}=\sum_{h=1}^{64} M_{kh}\exp^{-}(h-j/\sigma)^2 \; ; \; k,j=1,\ldots,64 \qquad (1a)$$

$$T'_{ij}=\sum_{k=1}^{64} S'_{kj}\exp^{-}(k-i/\sigma)^2 \; ; \; i,j=1,\ldots,64 \qquad (2a)$$

$$\Rightarrow GFVS=\left\{(i,j): \max_{ij} T'_{ij}\right\}.$$

This cortex model of the transform is the basis of the present work in CTT modeling at AFIT.

Diagrammatically, this gestalt process looks like this:

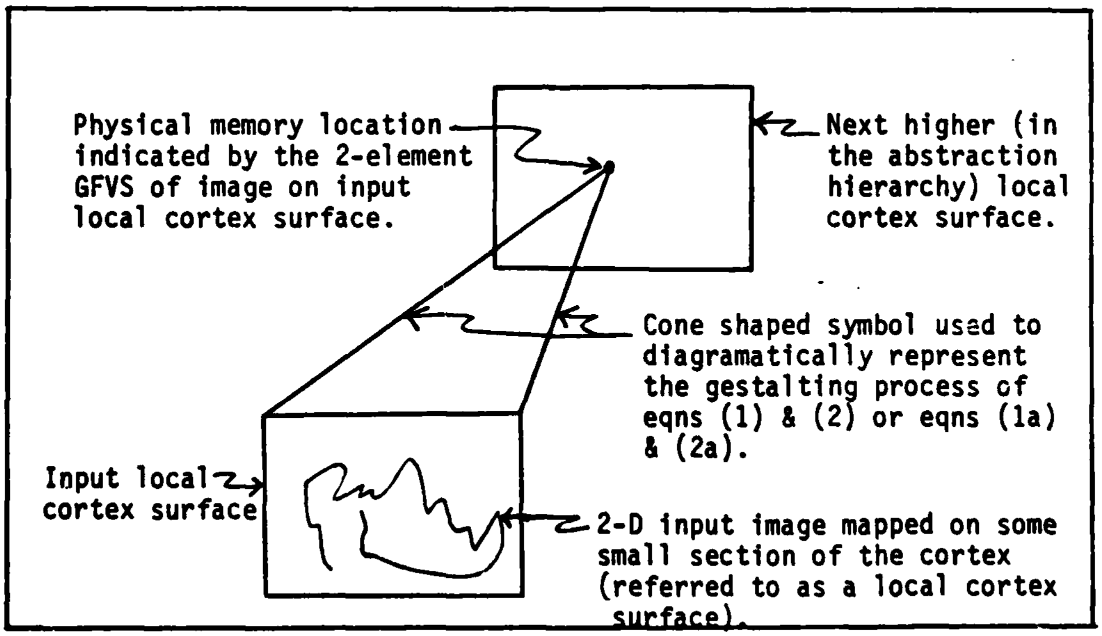


FIGURE 3.

Note that this architecture accounts for the phenomenon of direct memory access.

IV. EXPERIMENTAL RESULTS

Speech Recognition Results.

The following neurophysiologically suggested CTT partial model of the human speech recognition system was built:

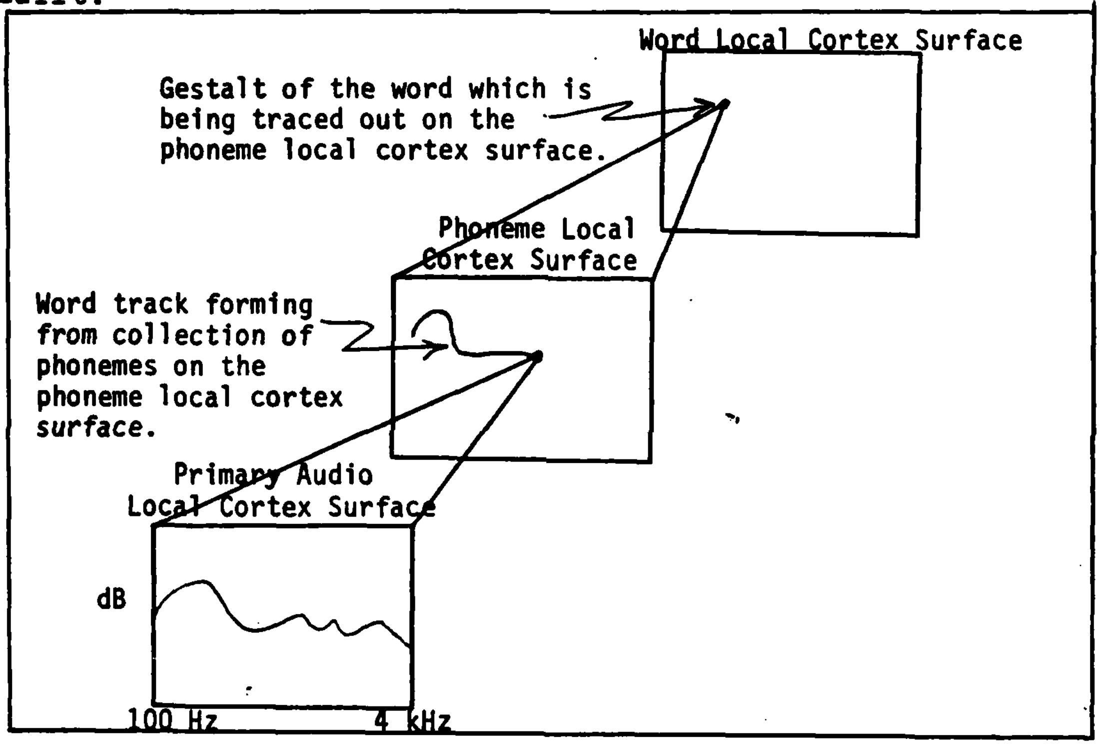


FIGURE 4.

The mapping of the vowels on the phoneme local cortex surface are shown in Figure 5. There is a startling similarity between this map and the Tragerian English Vowel alternation model. Upon a more detailed examination of the CTT audio research results, it appears that this CTT audio information processing architecture is capable of accounting for not only this psychological phenomenon of phonetic vowel alternation, but several other psychological phenomena, phoneme substitutions, deletions, and modifications, which commonly occur in connectted speech, as well as performing speaker dependant connected-word speech recognition. (It is speaker dependent due to the particular engineerng implementation of the acoustic preprocessing algorithms which were used for the model.)

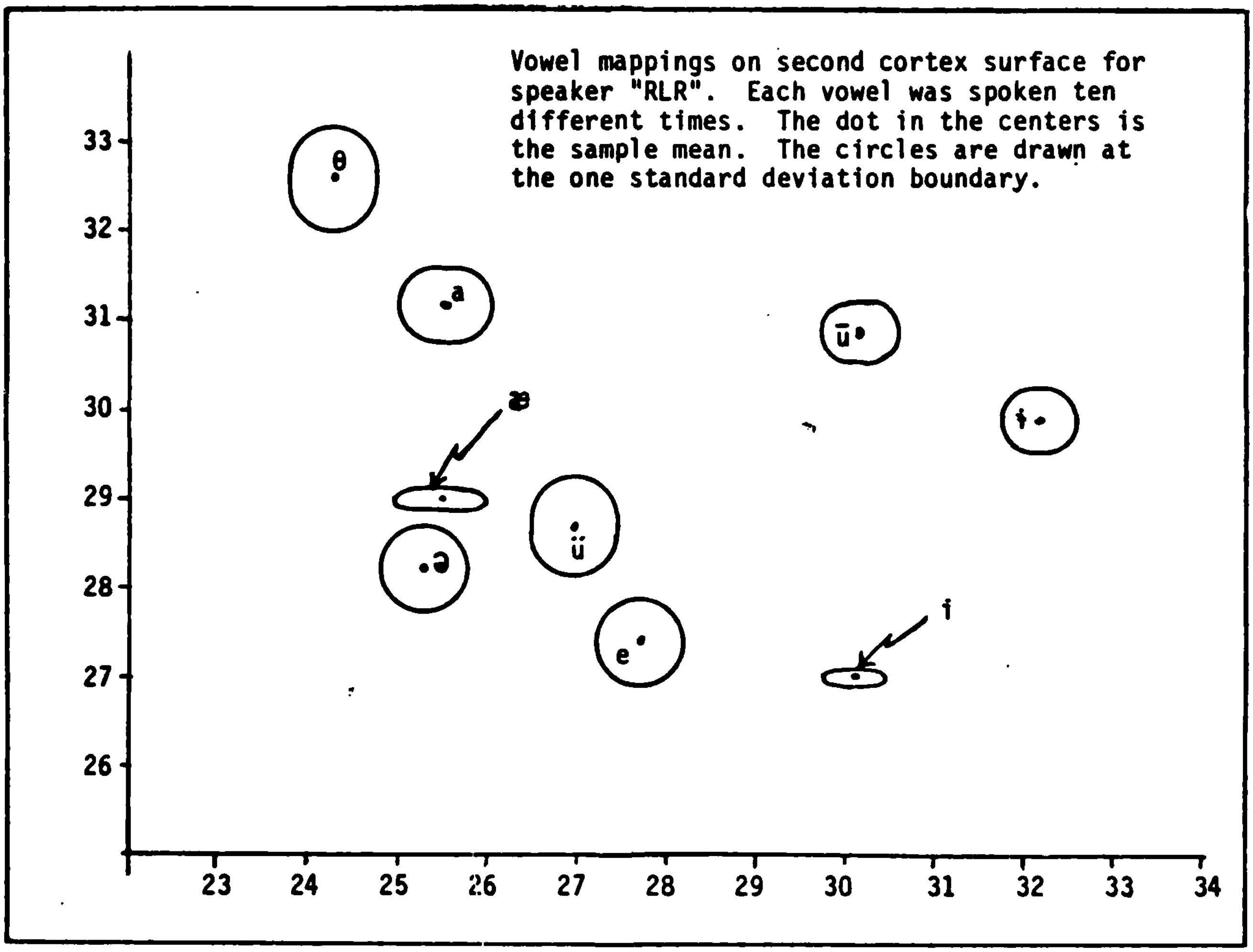

Vowel mappings on second cortex surface for speaker "RLR". Each vowel was spoken ten different times. The dot in the centers is the sample mean. The circles are drawn at the one standard deviation boundary.
33
32
31
30
29
28
27
26
23
24
25
26
27
28
29
30
31
32
33
34
θ
a
æ
ə
ü
e
ū
i

Figure 5

Image Recognition Results.

Russel has extended the original CTT research work into the visual image processing domain. The CTT architecture was implemented in the visual image processing domain and used to do human face recognition. The task of face recognition was chosen due to the apparent great difficulty previous conventional attempts have had in attempting to solve this problem. It was considered sufficiently difficult so as to provide a persuasive demonstration of the powerful advantages of a CTT approach.

CTT claims to be a generic model for sensory information analysis, regardless of the domain or entry level of abstraction. Russel decided to test this hypothesis by using the speech recognition program to do face recognition. He removed the 64 x 64 primary audio cortex map and inserted in its place a 64 x 64, sixteen gray level, digitized image of a human face. All the rest of the program remained unchanged. The preliminary results of this analysis, applied to five images each of sixteen different people, are shown mapped in Figure 6. The only bearded man, also partially balding, was significantly at one extreme of the spread. Those closest to him are partially balding. Two identical twins (separable by a first-time human observer -- but nevertheless admittedly quite similar in appearance) were classified by the CTT system as similar.

The preliminary indication of these results is that CTT is indeed a generic human classification architecture which produces psychologically similar results to that of a human.

Continued development, beyond the preliminary results shown here, has resulted in a robust CTT face recognition machine which has demonstrated high reliabilities for the proper identification of a new image input from a working knowledge base of approximately five prototypes each of twenty different people.

Preliminary performance measurements yield an accuracy of 91% for 22 pictures tested against a database of 20 people. (The two people who were not properly identified as first choice came in second and third, respectively.)

V. ABSTRACT HUMAN REASONING WITH A CTT ARCHITECTURE

When the cortex set-completion and sequence-completion mechanism hypothesized by Goldschlager are included with the CTT gestalt mechanisms, the result is an architecture which

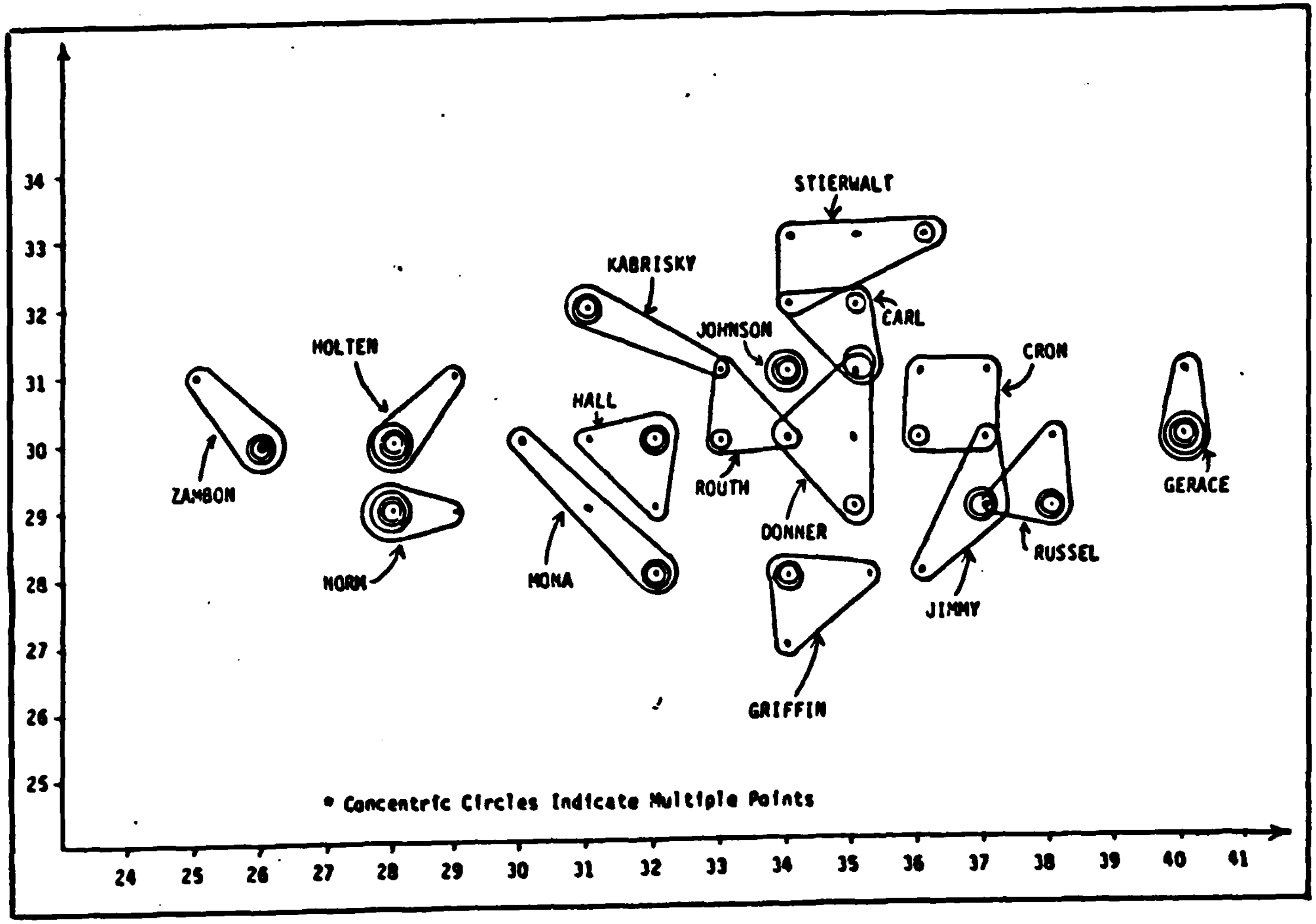


Figure 6 - Full Face Gestalt Plots

is sufficiently rich enough to disambiguate the sentence, "John shot the buck." Several previously observed phenomena characteristic of human natural language processing are accounted for by this model. A detailed analysis of this architecture reveals a structure sufficiently rich enough to account for the abstract reasoning behavior of the human brain.

Figures 7, 8 and 9 are presented as brief diagrammatical examples of the CTT structure necessary to disambiguate, "John shot the buck."

VI. LEARNING

In contrast to the conventional categories of learning found in the A.I. Literature, this work prefers to regroup learning into the following three categories which are (all three) accounted for by the CTT model: (1) learning in accordance with innate ability, (2) learning by multiple exposure to the same (or similar) cortex image, and (3) learning by attaching multiple associations to new observations.

VII. FUTURE DEVELOPMENT

In addition to any speculation as to the long term impact of CTT, some present and near term applications of CTT are suggested. It appears that it is now possible to build a real-time, connected-word, few hundred word speech recognizer. There also appears to be the promise of developing a low bit (100 bps) speech transmission system. Also, a sophisticated near real-time image recognition machine continues to be developed at AFIT.

BIBLIOGRAPHY:

For a complete bibliography, see the bibliography in:

ROUTH, RICHARD LEROY. Cortical Thought Theory: A Working Model of the Human Gestalt Mechanism. Ph.D. Dissertation AFIT/DS/EE/85-1. Air Force Institute of Technology: WPAFB, Ohio, July 1985.

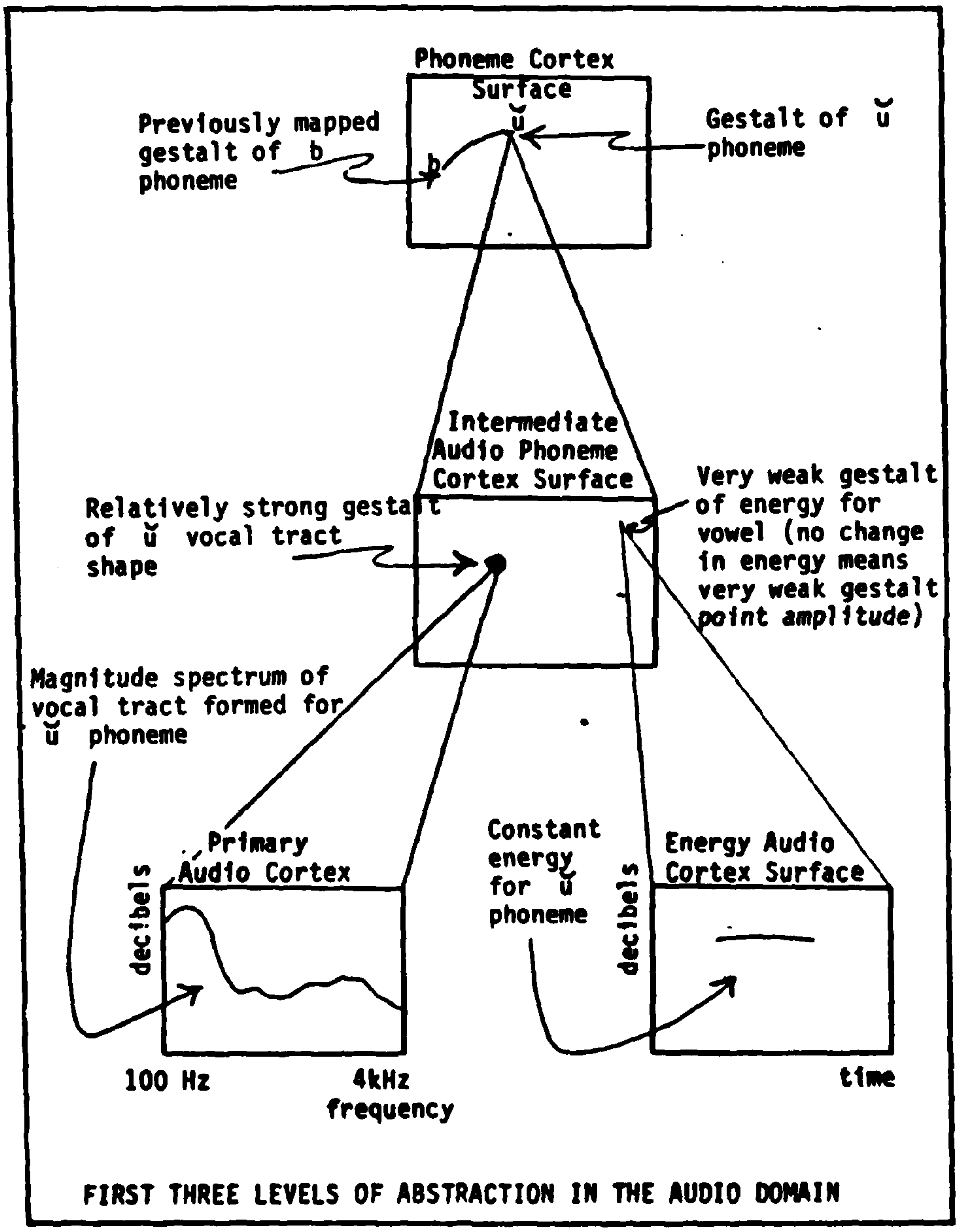
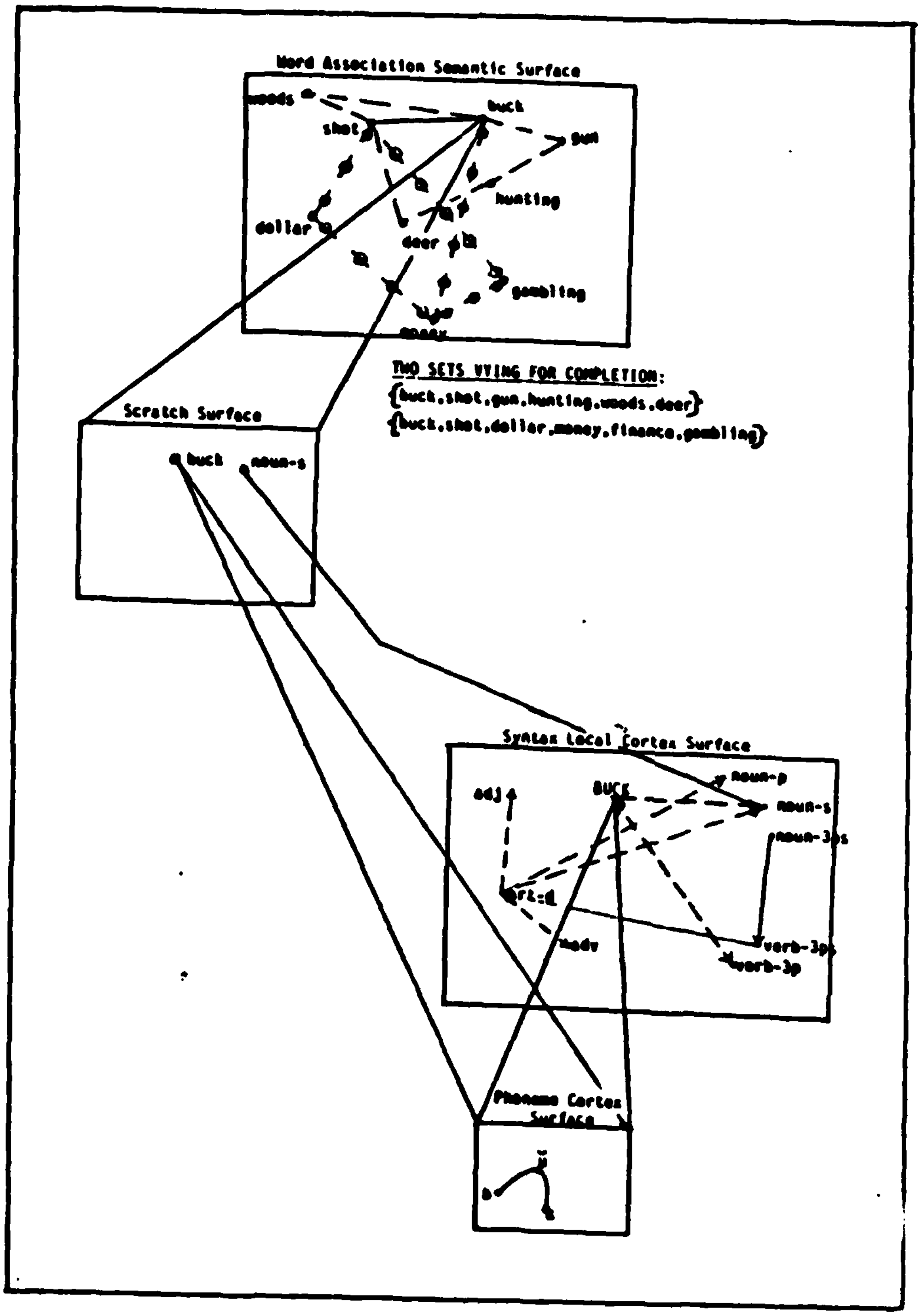
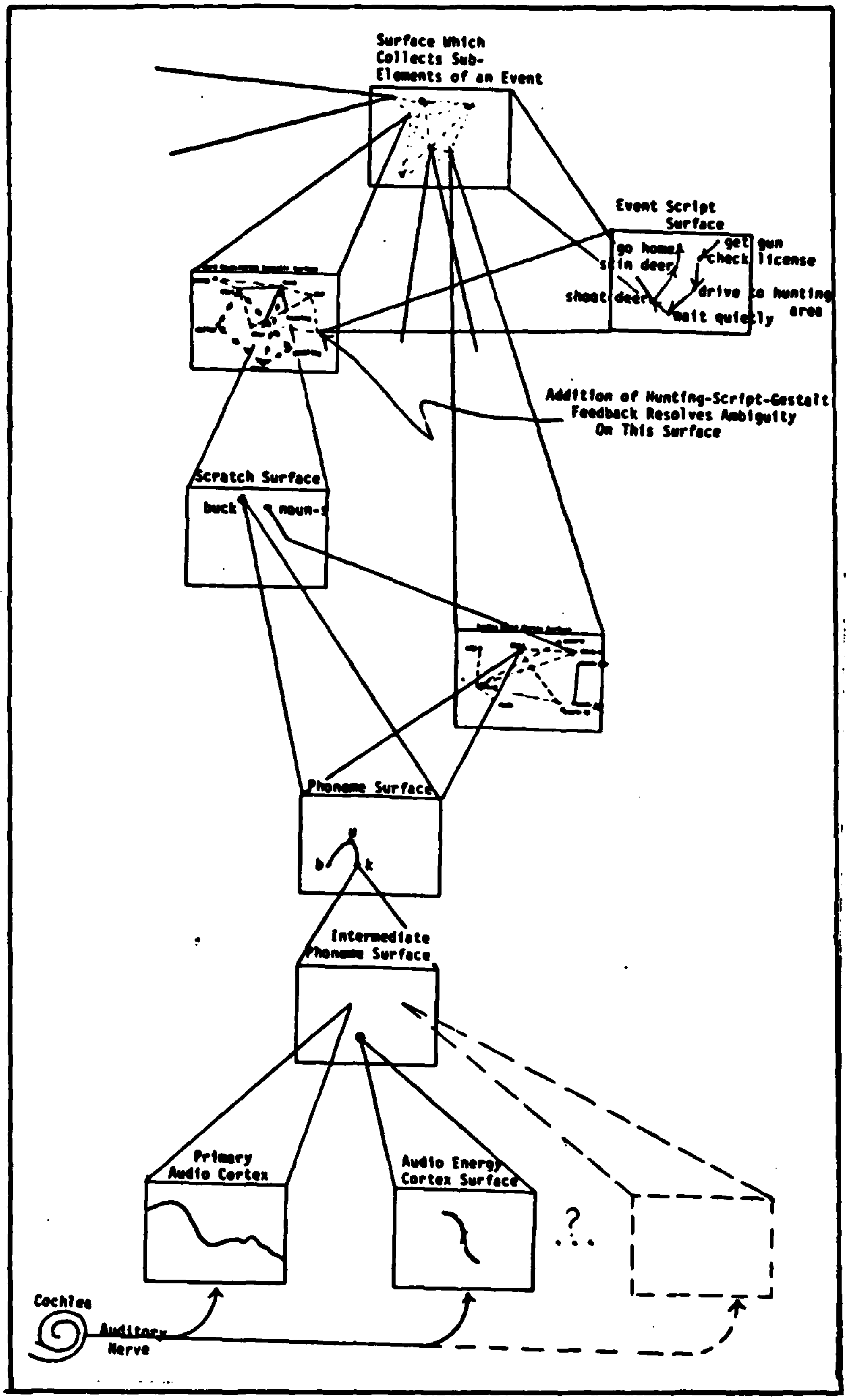

Phoneme Cortex Surface
Previously mapped gestalt of b phoneme
ŭ
b
Gestalt of ŭ phoneme
Intermediate Audio Phoneme Cortex Surface
Relatively strong gestalt of ŭ vocal tract shape
Very weak gestalt of energy for vowel (no change in energy means very weak gestalt point amplitude)
Magnitude spectrum of vocal tract formed for ŭ phoneme
Primary Audio Cortex
decibels
100 Hz
4kHz
frequency
Constant energy for ŭ phoneme
Energy Audio Cortex Surface
decibels
time
FIRST THREE LEVELS OF ABSTRACTION IN THE AUDIO DOMAIN

Figure 7



Word Association Semantic Surface
woods
buck
shot
gun
hunting
dollar
deer
gambling
money
TWO SETS VYING FOR COMPLETION:
{buck,shot,gun,hunting,woods,deer}
{buck,shot,dollar,money,finance,gambling}
Scratch Surface
buck
noun-s
Syntax Local Cortex Surface
adj
BUCK
noun-p
noun-s
noun-3ps
madv
verb-3ps
verb-3p
Phoneme Cortex Surface

Figure 8



Surface Which
Collects Sub-
Elements of an Event
Event Script
Surface
go home
get gun
skin deer
check license
shoot deer
drive to hunting
area
wait quietly
Addition of Hunting-Script-Gestalt
Feedback Resolves Ambiguity
On This Surface
Scratch Surface
buck
noun-s
Phoneme Surface
b
k
Intermediate
Phoneme Surface
Primary
Audio Cortex
Audio Energy
Cortex Surface
?
Cochlea
Auditory
Nerve

Figure 9

Appendix B

User's Guide for CTT Face Recognition System

This user's guide will cover four areas:

1) Calibration

2) Taking the picture of a face

3) Processing the facial image

4) Training the system and Recognizing a Face

CALIBRATION

A camera "calibration" check should be performed at the beginning of the session, to check primarily for proper camera warmup.

1) Studio Setup. Arrange the studio as shown in figure B-1.

2) Camera Settings. Turn on the camera using the red switch at the back. Remove the dusk cover from the lens. Set the controls as follows:

F-STOP: F5.6

FOCUS : 30 ft

ZOOM : 18mm

Let the camera warm up about 10 minutes before doing anything else.

3) Commands at NOVA terminal:

DIR NRUSSEL

PICTURE2

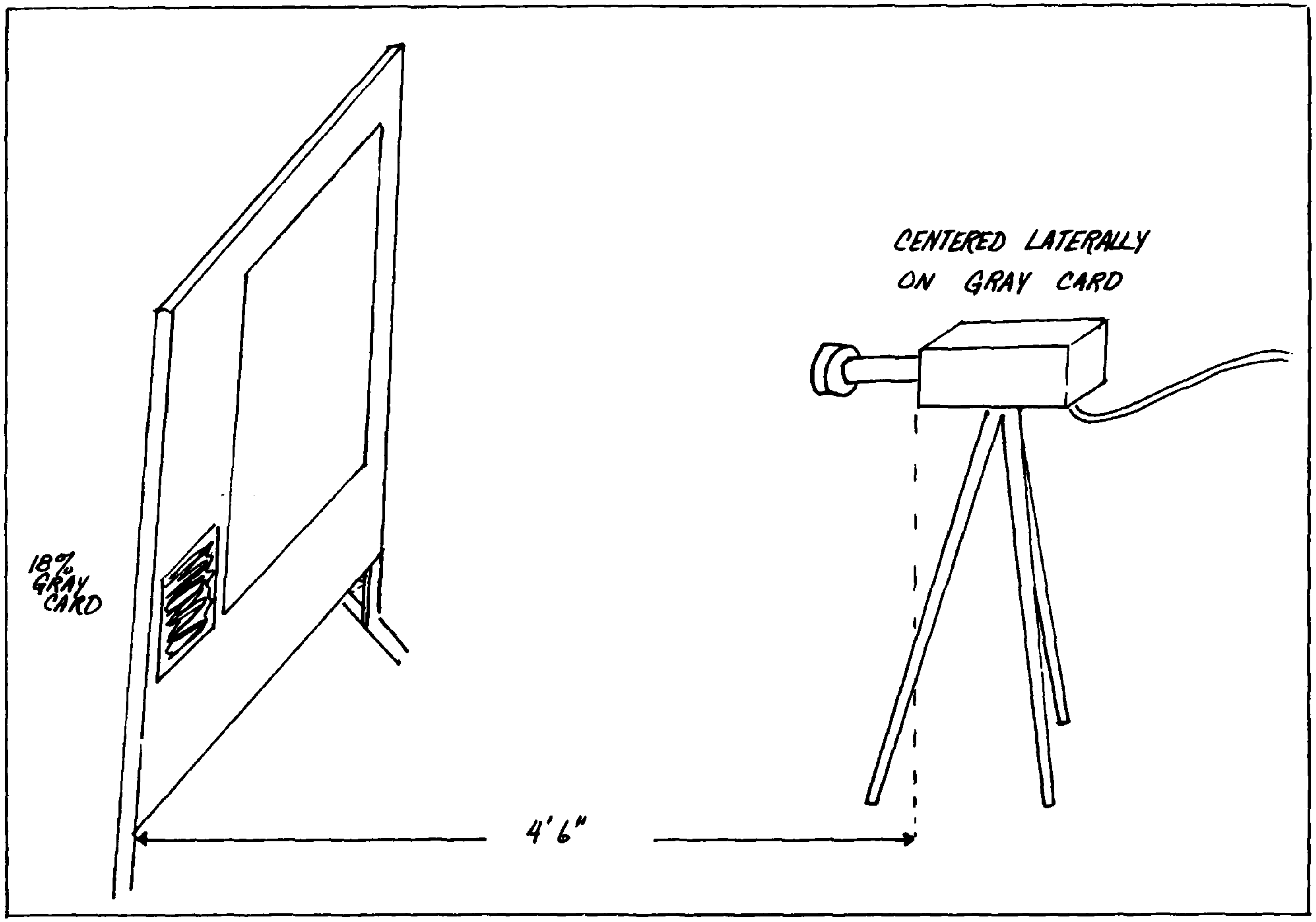


Figure B-1. Setup For Calibration (Side View)

The following menu is displayed on the NOVA terminal:

```
 * * * Cortical Thought Theory (CTT) Vision Processor * * *
                  by Robert L. Russel Jr.
   (Adapted from NOVA Sight Processor by James Holten III)

                * * * Keypad Menu * * *
-------------------------------------------------
       < FAST >                 < SLOW >
    1   2   3   4            5   6   7   8
-------------------------------------------------
    1--camera #1 on
    2--change rectangular window size, (now =64,64)
    3--crosshairs on
    4--camera #1 off
    5--menu of other options
    6--rectangular window
    7--box cursor on
    8--save current picture
    4 AND 5 -- terminate and exit to system

  Interactive video input control
```

When the menu is finally displayed on the terminal screen, and the terminal prompts with "Interactive video input control", then the user can enter further commands (see figure B-2.)

a) Hit the 1st key on the bottom row of the OCTEK keypad to select "Camera On".

b) Hit the 7th key on the bottom row of the OCTEK keypad to select "Box Cursor On".

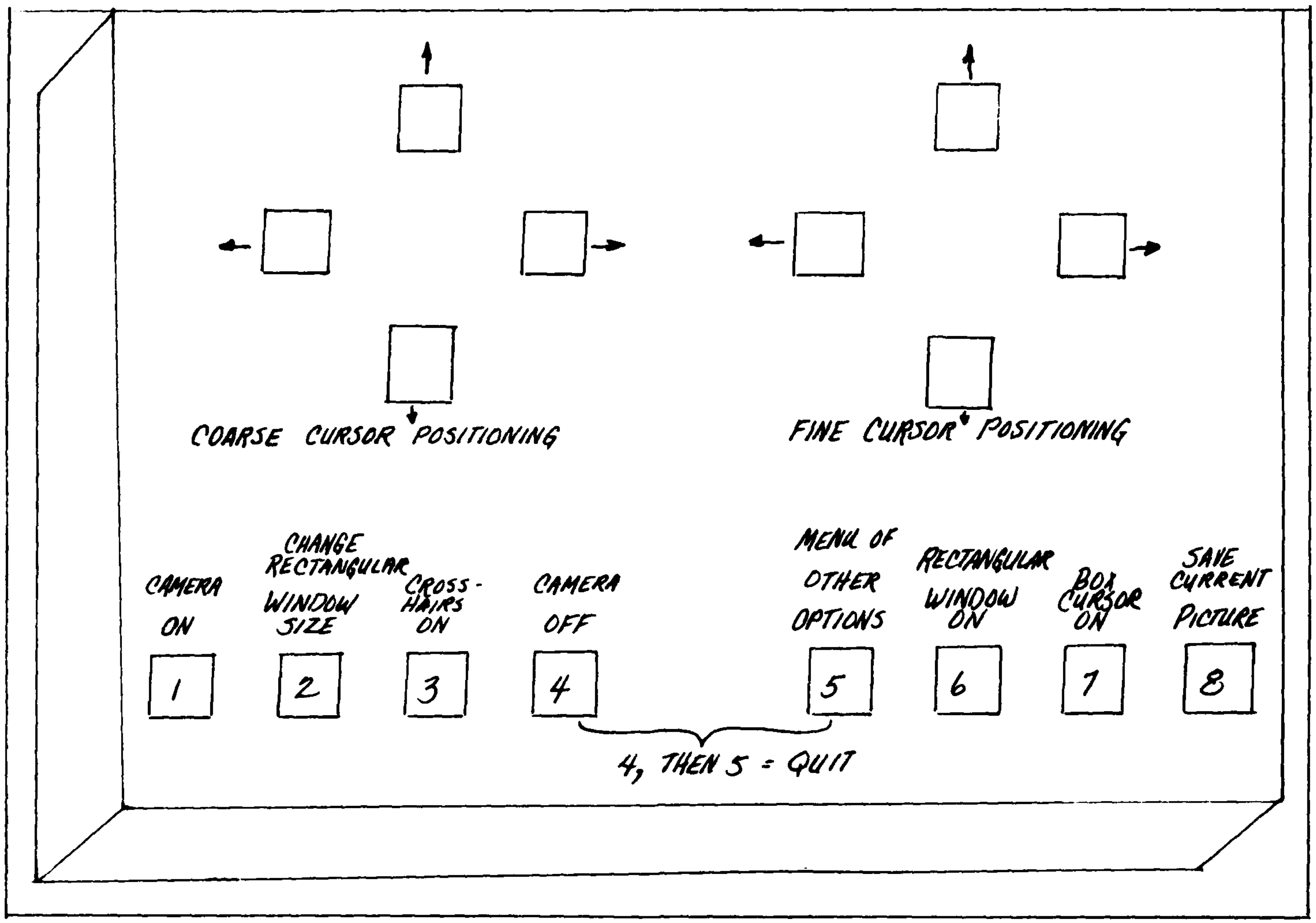


Figure B-2. OCTEK Keypad Controls for Program "PICTURE@"

c) Use the "Coarse" & "Fine" cursor positioning keys at the top of the OCTEK keypad, to position the box cursor over the Gray Card (see figure B-3.)

d) Adjust the zoom as necessary, so that the box cursor fits just within the gray card (see figure B-3.)

e) Hit the 4th key on the bottom row of the OCTEK keypad to select "Camera Off."

f) Push key #5 on the bottom row of the OCTEK keypad to select "Other Menu Items."

The following menu will appear on the NOVA terminal:

```
            * * * Optional Menu Items * * *

1 - Select Camera #1 (default value)
2 - Select Camera #2

3 - Retrieve a 64x64 file

4 - Set screen to WHITE
5 - Set screen to BLACK

6 - Negative/Positive Image (POSITIVE selected)
7 - Find Average Pixel value of area within box cursor

8 - Return to Previous Menu

        Choice:
```

g) Hit a "7" on the NOVA Keyboard (not the OCTEK keypad) to select "Find Average Pixel Value".

After about a second, the average pixel value will

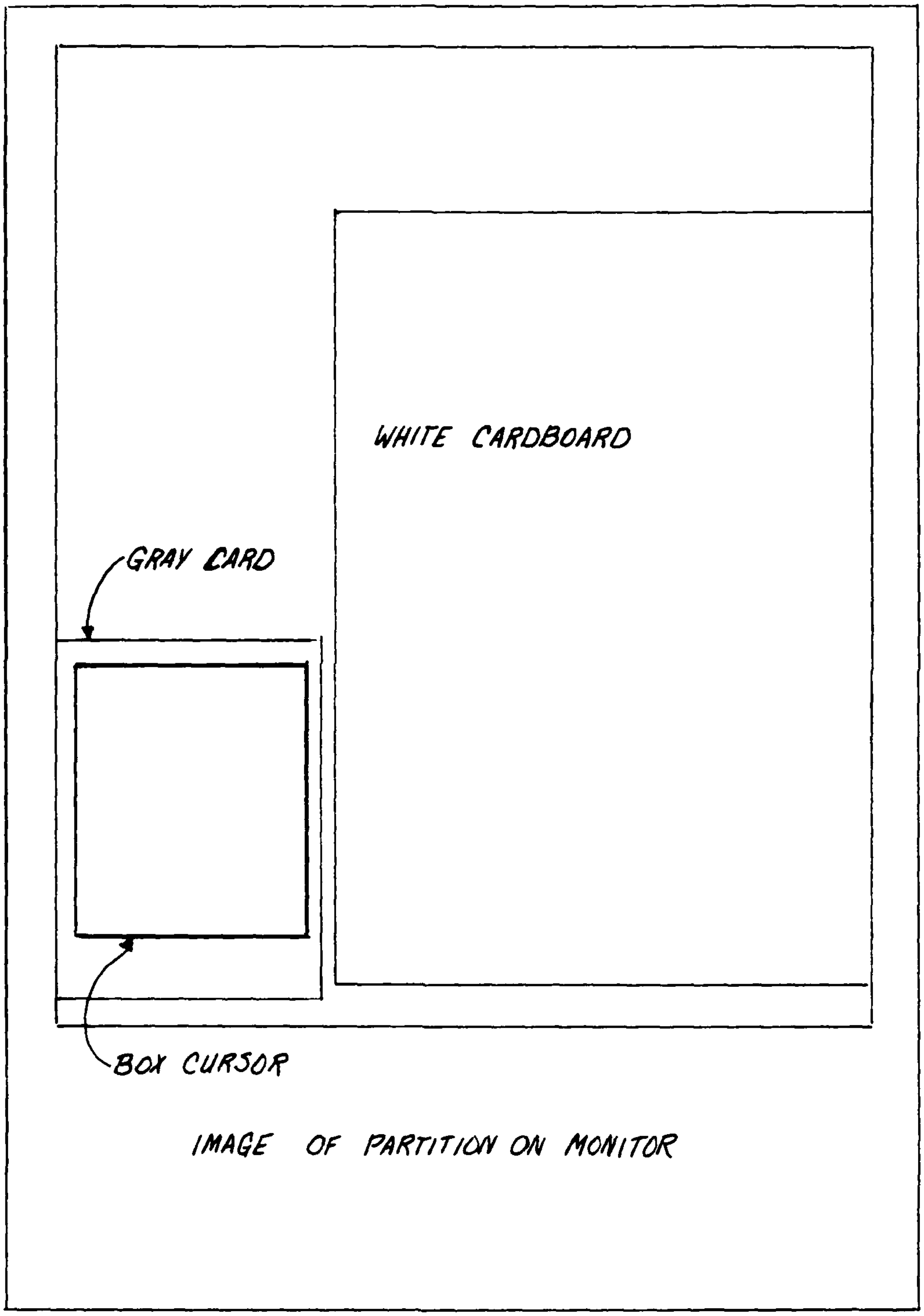


Figure B-3. Positioning of Box Cursor on Gray Card

be displayed on the terminal. For example,

===>> Average Pixel Value = 9.69

(This value should be between 9.0 and 10.0. If it is not, then the camera probably needs more warm-up. Wait a minute, then try again.)

TAKING A PICTURE

1) Studio Setup. Setup the studio as shown in figure B-4.

2) Camera Settings.

F-STOP: F8

FOCUS : 8 ft

ZOOM : Adjust as necessary (see below)

3) Commands at Nova Terminal:

a) The user should be using the program PICTURE2, as discussed under "Calibration."

b) Hit key #7 on the bottom row of the OCTEK keypad to select "Box Cursor On" (see figure B-2.)

c) Instruct user to sit up straight in chair, and look straight into camera.

d) Using the cursor positioning keys at the top of the keypad, move the box cursor so that it is centered over the face. Adjust the zoom and box cursor positioning until the top of the box cursor is on the top of the head,

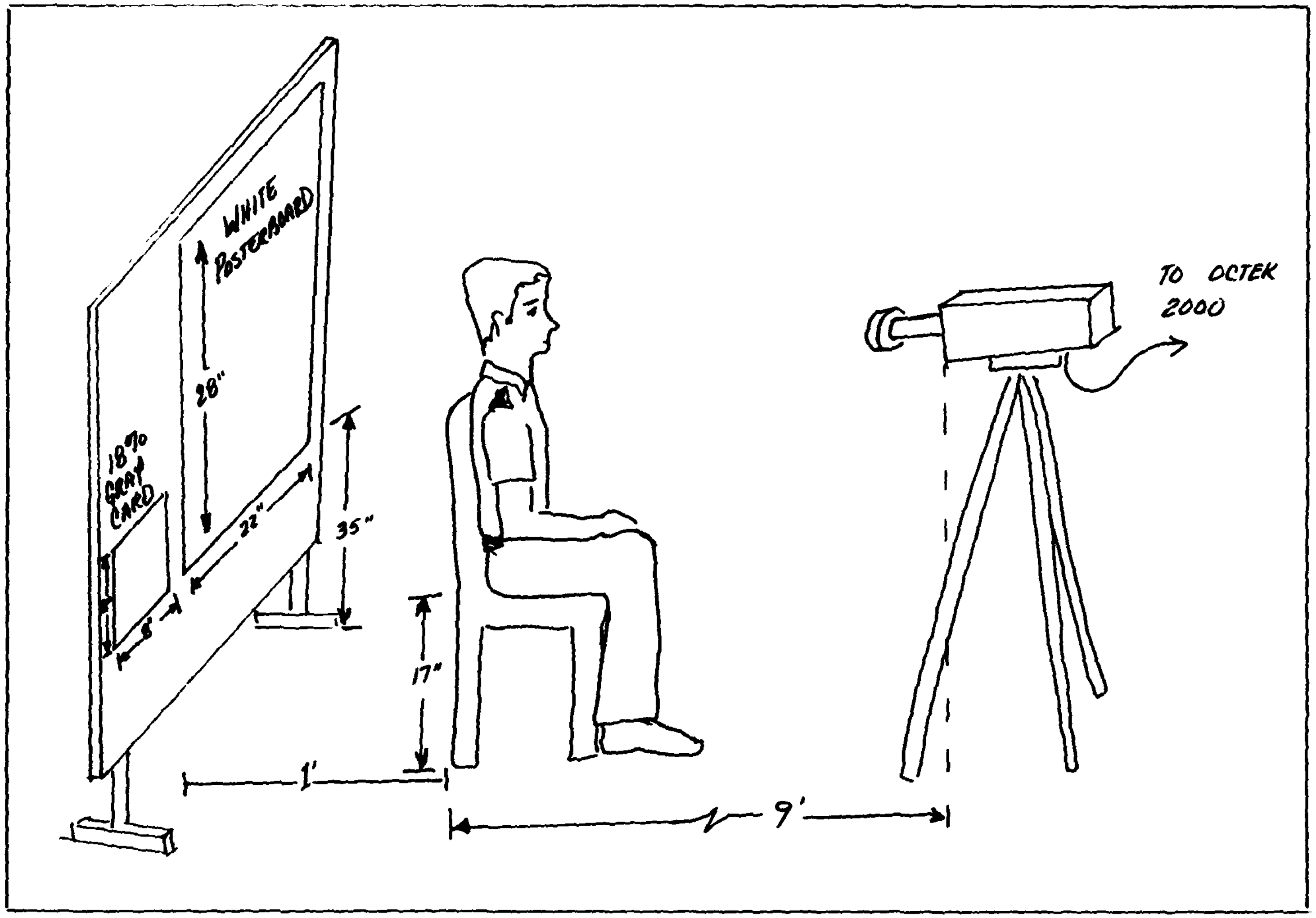


Figure B-4. Studio Setup For Taking A Picture

and the bottom of the box cursor is on the lowest light area on the chin (see figure B-5.)

e) When the user is looking directly at the camera, push key #4 on the OCTEK keypad, "Camera Off", to capture the picture.

HINT

Many subjects tend to turn their heads so they are not directly facing the camera (regardless of how you instruct them to position their heads.) To correct this problem, have them turn their heads about 10 degrees to one side, and slowly rotate their head to the other side. As they pass the correct position, take the picture.

f) Tell the subject they can now relax.

g) The box cursor must now be reduced to properly fit the head. Hit key #2 on the OCTEK keypad to select "Change Rectangular Window Size." See figures B-6 and B-7. Use the following criteria in making the adjustments:

(1) Left Side of Head. Use the keys on the upper left of the keypad to move the left side of the box to the left side of the hair on the head. (Ignore the ears. Also ignore hair which is below the ear level.)

(2) Top of Head. If necessary, move the

Figure B -5. Initial Zoom & Box Cursor Adjustment

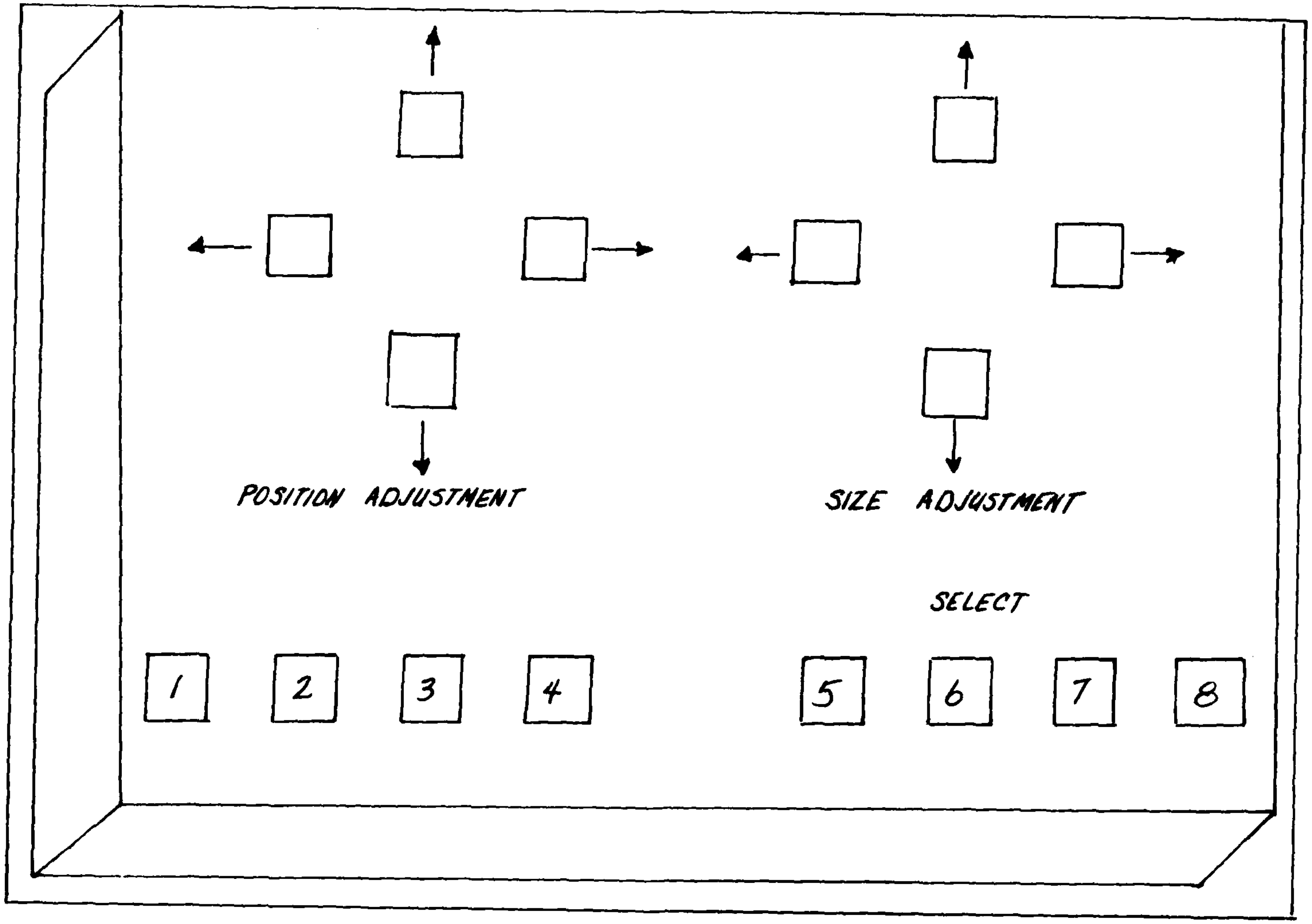


Figure B-6. OCTEK Keypad Controls for Adjusting Box cursor Size

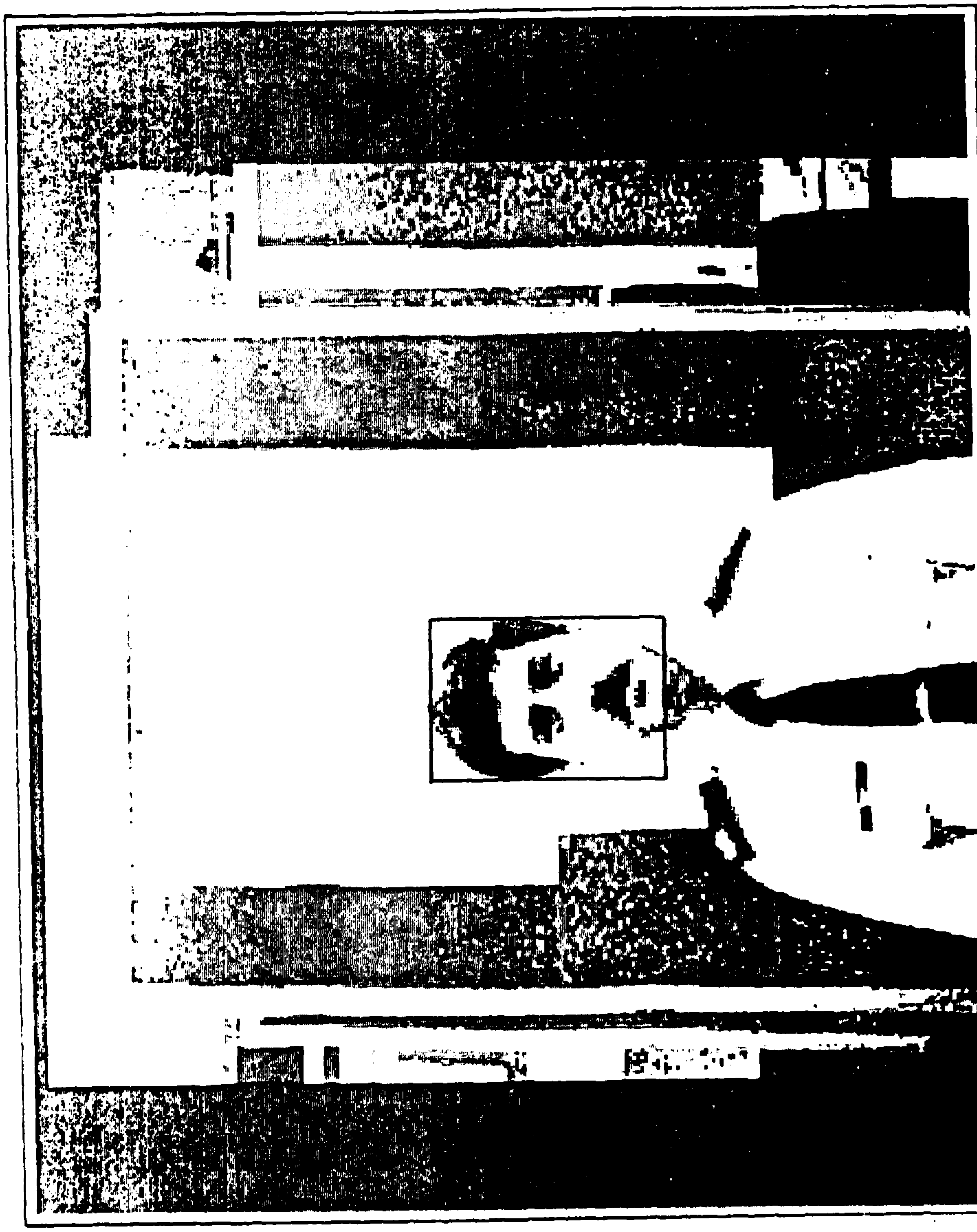


Figure B-7. Final Windowing with Adjustable Box Cursor

box cursor to properly fit the top of the head. Position the top of the box at the top of the general outline of the hair; ignore small tufts of hair.

(3) Right Side of Head. Using the left and right size adjustment keys at the upper right of the keypad, adjust the horizontal size of the box until it fits the right side of the head. Use the same criteria as for the left side for positioning.

(4) Lowest light area on chin. If necessary, use the upper & lower size adjustment keys to adjust the lower line on the box cursor to the lowest light area on the chin.

NOTE: If the vertical adjustments end up being more than a couple pixels, it is best to take the picture over.

(5) When done, hit key #6 on the lower row of the OCTEK keypad to return to the main menu.

h) Saving the image. Hit key #8 on the keypad to save the image. the program will respond with the following:

    Output file name:

Enter the filename at this point. The naming convention used on this system is as follows -- use up to the 1st 8

letters of the person's last name, then a number indicating which number picture it is in a series, and finally an extension of ".PI". For example, the name for the 23rd picture of Robert Russel would be:

```
RUSSEL23.PI
```

After entering the filename, the system displays the following menu:

```
Enter F-Stop used on camera:

    1 - F11
    2 - F8
    3 - F5.6

    4 - F4.0
    5 - Other
    6 - Unknown or Not Important

        Choice:
```

In this case, select #2 for F8. The program will then store the picture onto the disk. Then user can then go back and store another picture.

i) Disk Space. Be sure to allow enough disk space for all the pictures you add. If you get an error message when trying to store an image, then the usual problem is

lack of sufficient disk space. Go to one of the ECLIPSE terminals, and type the following:

```
DIR NRUSSEL
MOVE/R/V OHAIR -.PI
DELETE -.PI
```

The system will now move your recently-created files to the directory OHAIR, and delete them from the directory NRUSSEL.

i) To Quit. When the user is done acquiring pictures, he can quit by doing the following:

Hold down key #4 on bottom row of OCTEK keypad, and roll another finger over to key #5.

j) Backing up files. Back up your recently stored files by doing the following:

Go to an ECLIPSE terminal, and type:

```
DIR NRUSSEL
MOVE/R/V OHAIR -.PI
```

k) TURN

The user is now ready to process the pictures to extract their gestalt coordinates.

PROCESSING A PICTURE

Perform the following steps first on an ECLIPSE terminal:

1) Check file space and remove unnecessary picture files. The directory NRUSSEL only has enough room for about 5 - 10 picture files at a time. It is best to remove the files which are not being used, prior to loading needed files. Do this as follows:

```
DIR NRUSSEL
LIST/A -.PI
```

The picture files which are in this directory are now displayed. Remove filenames which will not be processed by typing the following for each file:

```
DELETE OHAIR filenamel.PI
```

2) Load files. Load the files you wish to process into the directory NRUSSEL, if they are not already there.

If the desired files are not present, load them by typing the following:

```
DIR OHAIR
MOVE/R/V NRUSSEL filenamel.PI
```

Repeat this until all the desired files are present.

To move groups of files, include the following wildcards in the filename as appropriate:

- `*` Any single letter
- `-` Any combination of letters

For instance, to move all the picture files for Routh to the directory NRUSSEL, type the following from an ECLIPSE terminal:

```
DIR OHAIR
MOVE/R/V NRUSSEL ROUTH -.PI
DISK
```

Make sure that there is at least a value of 300 showing for available disk space on the resulting display. If not, then the subsequent program will probably run out of room. If

more space is needed, some of the picture files have to be deleted. This is done as follows:

```
DIR NRUSSEL
DELETE FILENAME.PI
DISK
```

Repeat this until there is sufficient space available. The deleted files will have to be loaded after the presently loaded files are processed and deleted.

STARTING THE ECLIPSE GESTALT PROCESSOR. To start the gestalt processing program on the ECLIPSE computer, type the following on an ECLIPSE terminal:

```
DIR ERUSSEL
RUNFACE
```

The system will respond with the following:

```
MOVING TO NRUSSEL
Deleting Excess Files

* * * CORTRAN16 Gestalt Processor Program * * *

Do you want the results sent to the printer?
(1=Yes,2=No):
```

At this point the user usually would pick choice #1. The system responds with:

* * * READY TO PROCESS PICTURE DATA * * *

The user is done for now with the ECLIPSE terminal, and should move to the NOVA terminal.

PROCESSING ON NOVA TERMINAL

The user begins by typing the following:

FACE

The system responds as follows:

* * CORTICAL THOUGHT THEORY (CTT) FACE RECOGNITION SYSTEM * *

Please enter the name of the file you want to process:

The user should enter the filename at this point. The system responds with:

* * * STORING FILENAME * * *

Checking file for F-Stop Information

* * * Retrieving Original File * * *

The system now clears the monitor screen, and displays the requested facial image on the monitor (see figure B-8.) Next, it performs an initial contrast expansion on the image, displaying it in the top center of the screen (see figure B-9.) Then it calculates the feature locations on

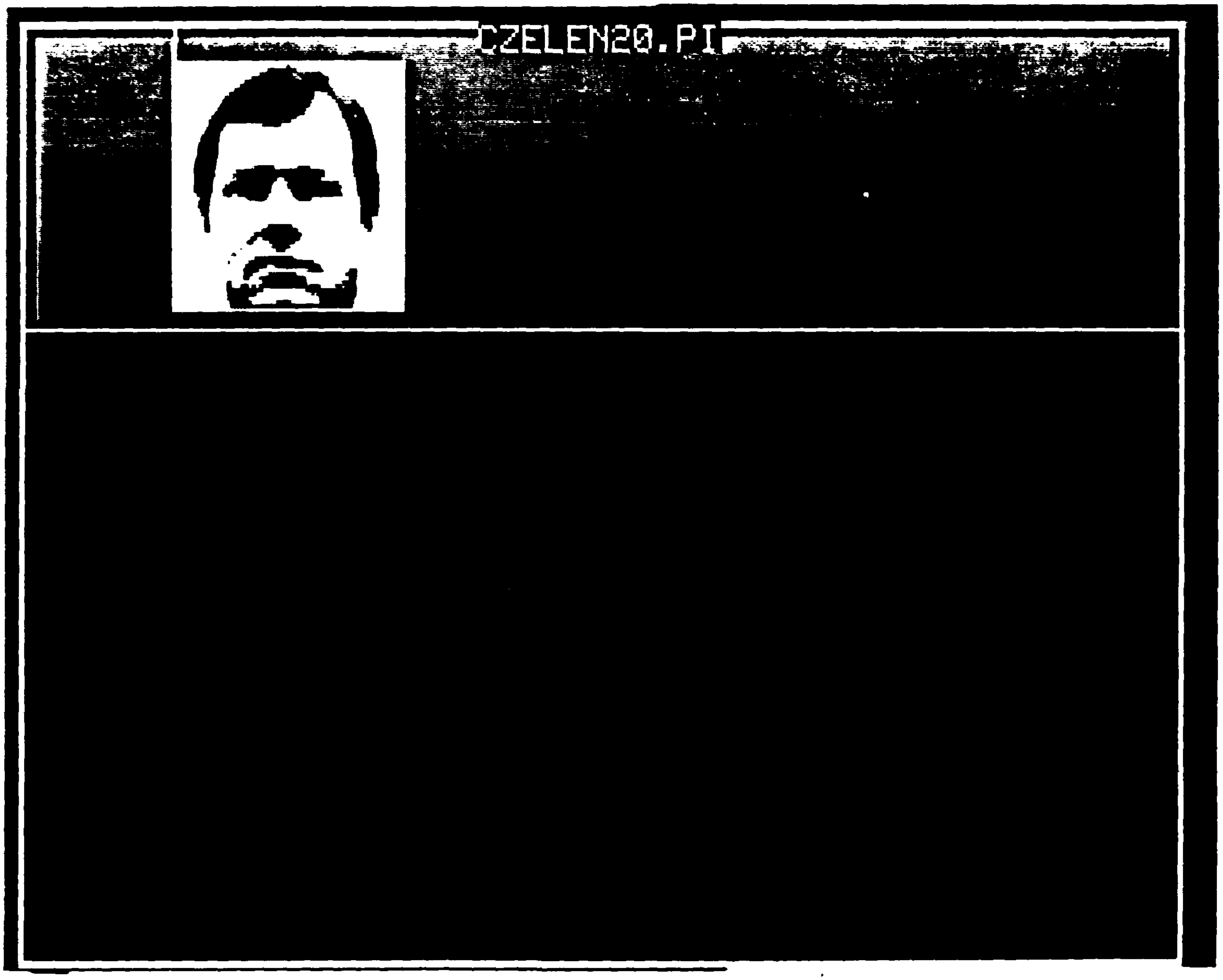


Figure B-8. Display of Original Image

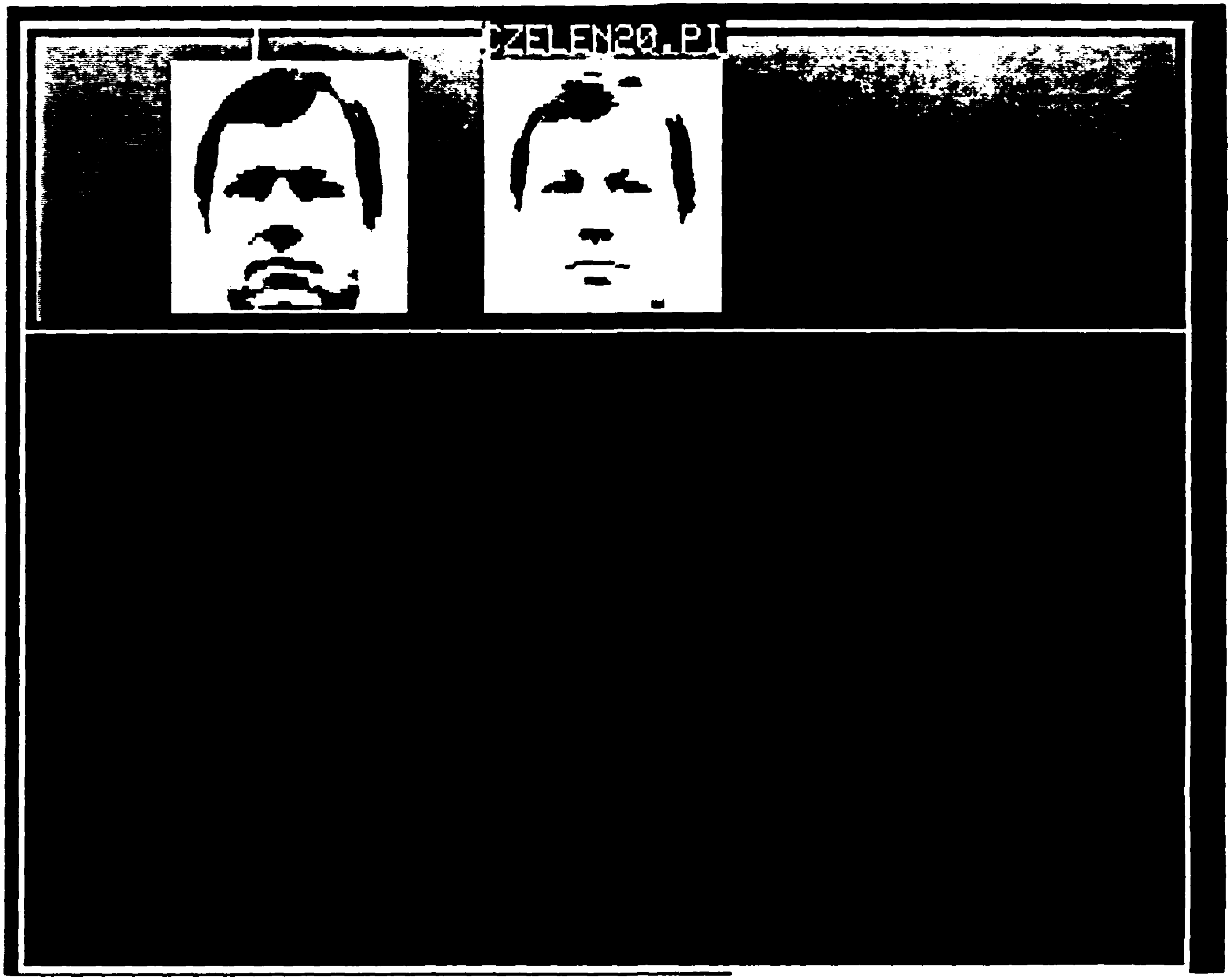


Figure B-9. Display of Initial Contrast-Expanded Image

the face, and displays them as lines on the image, at the upper right of the screen (see figure B-10.)

The user is now given the opportunity to change the feature locations.

```
Do you want to change any of the locations displayed?

    1 - Yes
    2 - No

        Choice: 1


  * * * Adjust Feature Positions * * *

 1 - Top of Head
 2 - Top of Eyes
 3 - Bottom of Eyes
 4 - Top of Nose
 5 - Center of Mouth
 6 - Lowest light area on Chin

 7 - Left side of Head
 8 - Right side of Head

 9 - Left side of Eyes
10 - Right side of Eyes

11 - Center of Eyes

50 - Return to Main Program

          Choice: 2
```

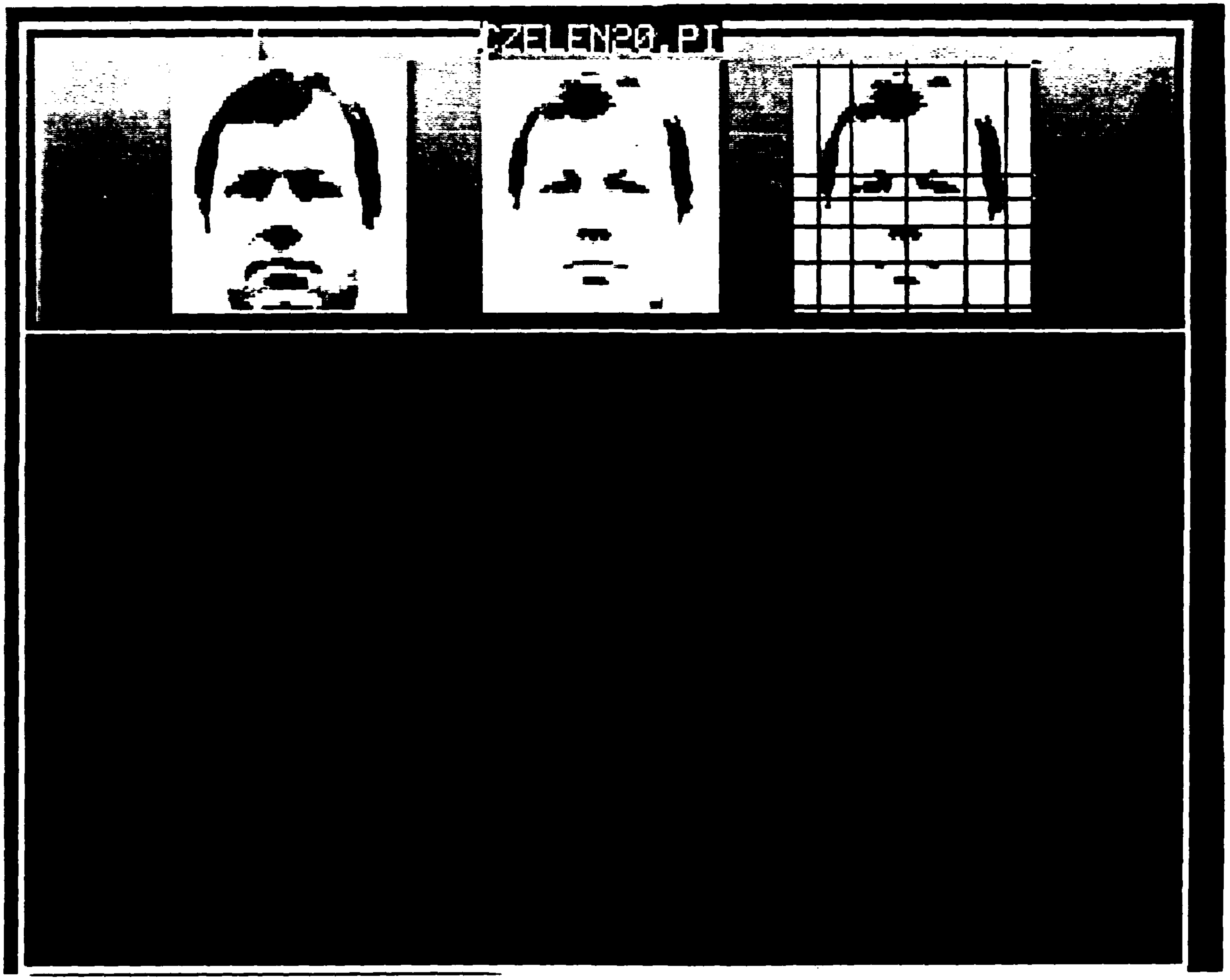


Figure B -10. Display of Feature Locations on Face

Suppose the user chooses to change the location of the top of the eyes, as shown above. The system will display crosshairs on the top center image, with the horizontal bar (in this case) located at the location calculated for the top of the eyes (See figure B-11). (If a vertically-aligned feature is chosen, such as the side of the head, the vertical bar on the crosshairs is then used to show this feature.)

To adjust the location, use the four keys on the upper right of the OCTEK keypad (see figure B-12). When done, hit the key on the far right on the lower row.)

The following criteria have been used in defining the locations of features:

1) Top of Head -- top of general outline of hair, independent of small tufts of hair here and there.

2) Sides of head -- the outside contour of the hair at about the ear level. Does not take into account hair which may curl out toward the bottom of the head.

3) Top of eyes -- measured as the top of the eyebrows.

4) Bottom of eyes -- the bottom of the shadow formed by the eye sockets.

5) Top of nose -- top of the dark area formed bythe nostrils when the image is contrast expanded. (the rest of the image will not show up in the final processed image.)

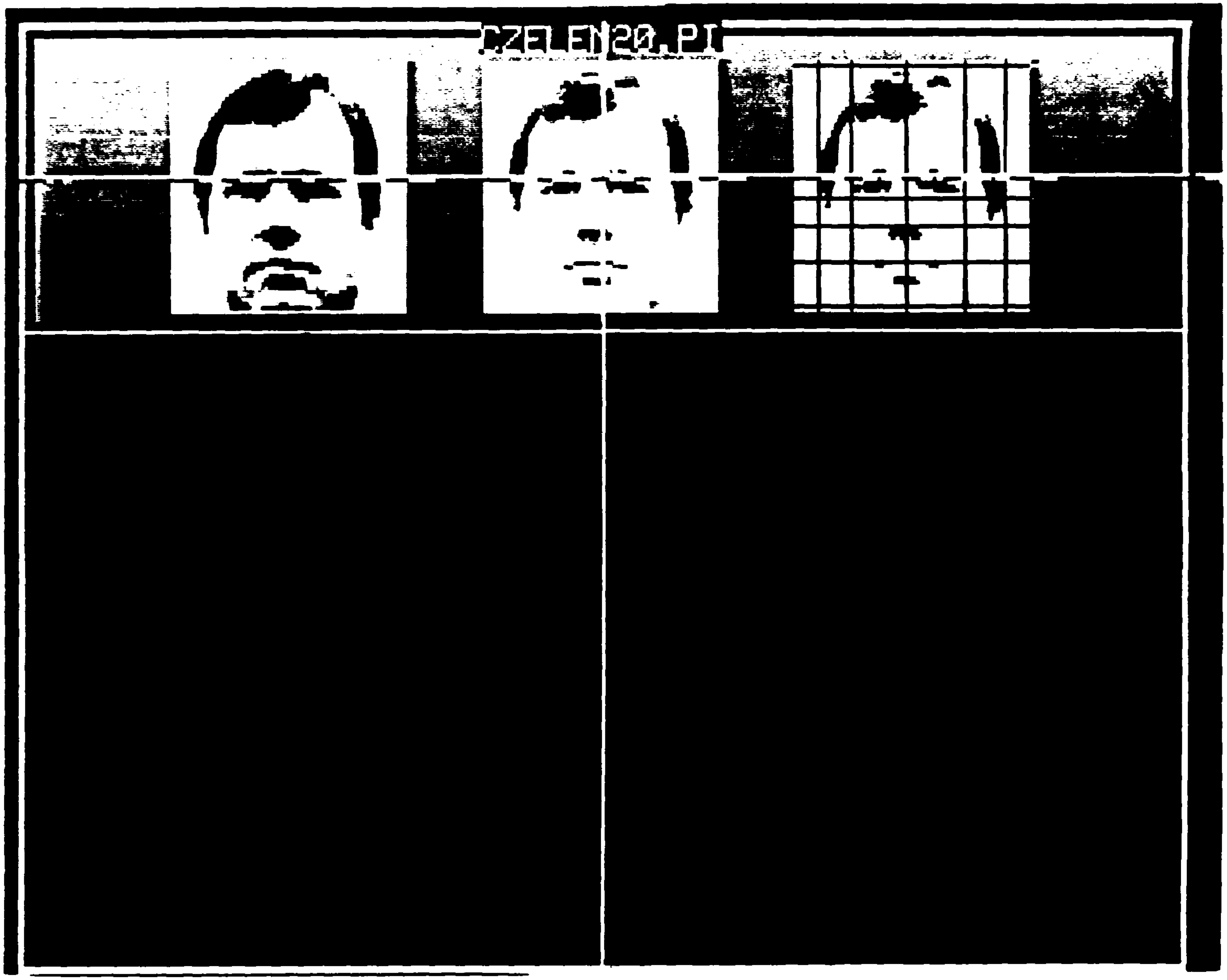


Figure B-11. User Adjustment of Feature Locations

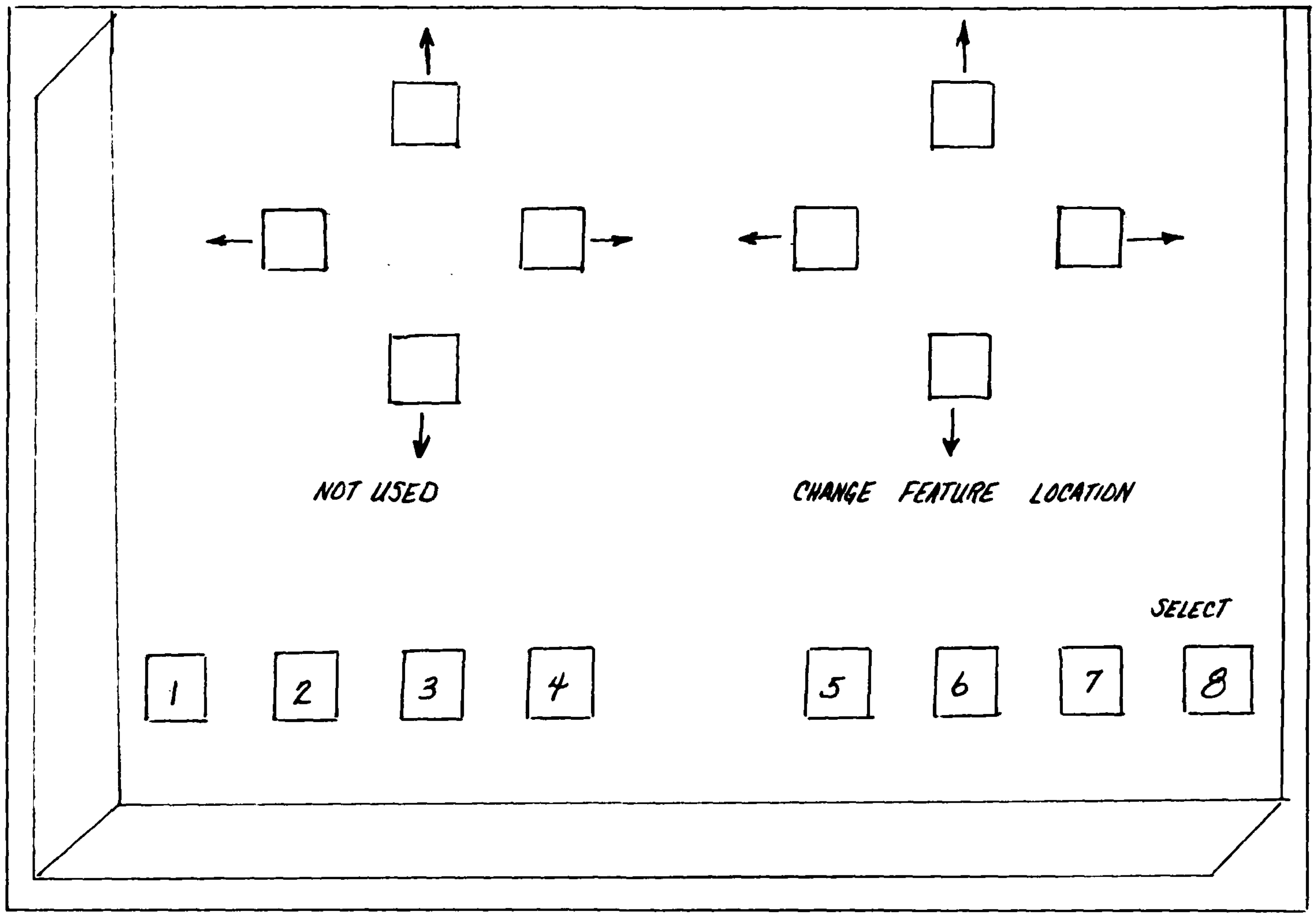


Figure B-12. OCTEK Keypad Controls for Adjusting Feature Locations

--- NOTE ---

Pay particular attention to the locations for the bottom of the yes and the top of the nose. If these are misadjusted, the final contrast expansion process may produce an undesirable image.

----

6) Center of mouth -- same as bottom of upper lip.

7) Bottom of chin -- the line formed by the lowest light area on the bottom of the chin . This is not necessarily the end of the chin itself, as the actual end of the chin is usually in shadow.

8) Left & right sides of eyes -- outside edge of eyes (not eyebrows).

The terminal gives directions for changing a feature location, as follows:

Use top RIGHT buttons to adjust the feature's location.

To ENTER this location, hit RIGHTMOST Button on BOTTOM of keypad.

The user now gets prompted for any additional changes.

```
Do you want to change any of the locations displayed?

    1 - Yes
    2 - No

        Choice: 2
```

Note: If at any further point in the processing, the user realizes that a feature location was incorrectly adjusted, he should do the following:

CTRL A

NEWFEAT

The system will now calculate a final contrast-expanded image, which may or may not look like initial contrast-expanded image. Next, the system extracts sub-images from the face, storing them to disk and displaying them on the screen (see figure B-13.) While the NOVA is extracting the images, the ECLIPSE begins calculating the Gestalt coordinates. These are then

Figure B-13. Extraction of Sub-Images from Face

displayed on the monitor above the picture to which they belong (see figure B-14.)

Now the user is given the option to print the monitor image on the ECLIPSE printer.

```
Do you want to print this image? (1=Yes, 2=No): 1
  Outputting File to Disk, Please Wait...
The File is Saved.  Now printing...
```

The user is now given the option to save the processed data as a record in the "Processed Picture" database.

```
Do you want to store this record? (1=Yes, 2=No): 1
Please enter the ID number of this person:
     1 - Enter ID number
     2 - Add a New Name
```

Figure B-14. Display of Gestalt Coordinate Values

```
3 - Use last ID number in system
       (LAST ID NUMBER = 12     Capt David King)

* * * Other Options * * *
9 - Look at or Edit Previous Records

       Choice:
```

If the user wants to store the record, the computer needs to know the identity of the person. If the user knows the ID number already, he can enter it directly. If he does not, he can look at a list of records by choosing option "9". If it is new, he can add it by selecting option #2. Finally, the system lets the user use the last ID number which was entered for a previous record. Each option is explained in more detail as follows:

Enter an ID Number

```
Enter ID number: 4
        ID Number = 4    Capt James R. Holten III

Please Choose an Option:

    1 - This is correct
    2 - Try Again
   -1 - Return to Main Menu
```

Choice: 1

At this point the system is happy, and will continue on to the new program.

Add a New Name

Enter the full name (with title, if desired):

<<=== You can enter up to here

Is this correct? (1=Yes, 2=No): " 1

* * * Updating Data Base * * *

The system will now store this ID number value as the user's choice, and continue to the next program.

Use Last ID Number in System

This choice will make the system use the last ID number stored in the system, if there was one. This last number, and the name of the person, is displayed below option 3 as follows:

```
3 - Use last ID number in system
       (LAST ID NUMBER = 12     Capt David King)
```

Once this selection is made, the system uses this ID number, and goes to the next program.

At this point, the user can select whether this record is to be Selected for Training, or Not Selected for Training. If Selected for Training, the record will be used to build the Recognizer Database. If Not Selected for Training, the record will reside in the Processed Records Database, but not be included in the Recognizer Database.

Of course, the training selection status can be readily changed at any time through the program MAIN on the ECLIPSE computer.

```
    Do you want the system to train with this record?
       (1=Yes, 2=No): 1
```

The processing is now completed. The system will now display the amount of file space left on NRUSSEL. This is done because the file space rapidly decreases with the more picture files stored in the directory. The user should keep at least a value of 300 for Available File Space to give the system sufficient room for processing. This is most easily done by only keeping 5 - 10 picture files at a time in NRUSSEL. The system deletes files which it creates itself and it no longer needs, so the user need not worry about system-generated files.

The space left on NRUSSEL is:

Space Available = 463   Space Used = 10000
Contiguous Space = 10

* * * Face Recognition Routine Completed * * *

When done processing faces, type QUIT.

FINISHING PROCESSING. When the user types QUIT, the system creates a flag file, telling the ECLIPSE computer to terminate the program which is processing Gestalt coordinates, moves picture files back into the picture file archive in the directory OHAIR, and starts the program MAIN on the ECLIPSE. MAIN allows the user to train the system for recognition, and to identify a person. At this point, the user should move to a ECLIPSE computer terminal for any further actions.

TRAINING FOR AND RECOGNIZING A PERSON

When the user types "QUIT" at the NOVA terminal, CORTRAN16 will terminate, the system will automatically delete excess files and move picture files back to the directory OHAIR, and then the program MAIN will begin on the ECLIPSE computer. The user can also invoke MAIN from scratch by typing the following at an ECLIPSE terminal:

```
DIR ERUSSEL
MAIN
```

This program will be used to build the Recognizer Database, and calculate the identity of a person.

TRAINING THE DATABASE

In order to train the database for a person, three steps must be taken:

1) Process & store records for at least 5 pictures of the person to be trained.

2) Make sure these records are flagged for training, either when they are stored, or after they are stored though menu item #2, "Select/Deselect Records for Training."

3) Train Recognition Database with Selected Records

by selecting item #3.

These steps are described below.

Step 1 -- Process & store 5 pictures. In order to provide the system a representative view of slight changes in expression, the subjects are asked to vary one or more of the following from one picture to another:

1) Comb hair slightly different (especially hair in front of head)

2) Slight smile or slight frown

3) Slight squint

Processing and storing of these pictures is done as described in the previous section.

Step 2 -- Flag records for training. The easiest point to do this is while processing the pictures, as the system will ask whether or not the records will be used for training. The second way to do this is through the program MAIN on the ECLIPSE. Menu item #2 is "Select/Deselect Records for Training" (see below):

```
* * * CTT Face Recognition System * * *

          - - - Main Menu - - -

 1 - List records in Main Database

 2 - Select/Deselect Records for Training
 3 - Train Recognition Database with Selected Records
```

```
4 - Load a Record from Processed Picture Database for
        Recognition
5 - Identify a Person

6 - Look at Maintenance Menu
-1 - Quit

        Choice:
```

At this point the user can manipulate the data from the processed pictures. Each of these functions will be explained below.

Menu item #2 is used as follows:

```
* * * Select Records for Training * * *

Do you want:  1 - Single Records, or  2 - Range of
                  Records?
```

For Single Records:

```
Which record do you wish to access?  120

Record:  120   ID Num = 12   Capt Cheryl Nostrand
    * * * SELECTED for Training * * *

Is this the correct record?  (1=Yes, 2=No):  1

Do you wish to do another single record? (1=Yes,2=No):
```

For Range of Records:

```
Enter initial record number:  1
Enter last record number:  10

Which would you like?

    1 - Select for Training
    2 - De-Select for Training

        Choice:  1

Do you wish to do another range of records?
        (1=Yes,2=No): 2
```

Once the user has completed selecting which records in the Processed Picture Database will be used to train the system, these records can be processed for training by selecting menu option #3 in the main menu, "Train Recognition Database with Selected Records."

Training the Recognition Database

```
* * * Calculating Face Recognition Database Statistics * * *

    Update statistics on file? (1=Yes,2=No):

(Choosing Yes retrains the database.  Choosing No only
calculates the statistics, but does not touch the database.)

    Print Results? (1=Yes,2=No):

(An example of the printed output is shown in Table B-1.)
```

* * * FACE RECOGNITION DATABASE -- STATISTICS CALCULATIONS * * *

(The smallest standard deviation is defined to be 0.5, in order to take care of discretation error.)

Date: 11/27/85   Time: 17:28

* * * CALCULATIONS FOR WINDOW 1 * * *

* * * STATISTICS FOR ID NUMBER 1, CAPT RON SMALL * * *

Total Number of Points in Database = 8
X Standard Deviation = .60
Y Standard Deviation = 1.66
Average X Value = 12.1
Average Y Value = 46.0

Minimum X Distance = 11
Maximum X Distance = 13
Minimum Y Distance = 44
Maximum Y Distance = 48

* * * STATISTICS FOR ID NUMBER 2, CAPT BOB RUSSEL * * *

Total Number of Points in Database = 9
X Standard Deviation = .67
Y Standard Deviation = .50
Average X Value = 14.0
Average Y Value = 47.0

Minimum X Distance = 13
Maximum X Distance = 15
Minimum Y Distance = 46
Maximum Y Distance = 48

* * * STATISTICS FOR ID NUMBER 3, CAPT MAX HALL * * *

Total Number of Points in Database = 10
X Standard Deviation = .50
Y Standard Deviation = .92
Average X Value = 14.0
Average Y Value = 40.4

Table B-1. Statistics Calculation (Part 1)

****

* * * SUMMARY OF WINDOW PERFORMANCES * * *

Window Number	X Perf.	Y Perf.	Figure of Merit
1	2.13	8.41	8.67
2	3.20	7.53	8.18
3	1.91	3.88	4.32
4	1.80	3.82	4.22
5	2.13	2.41	3.22
6	3.40	7.78	8.49

****

Table B-1. Statistics Calculation (part 2)

Please Wait...

* * * Statistical Calculations Done * * *

At this point, the system is ready to recognize a person.

Recognizing a Person

To recognize a person, the following must be done:

1) Load the gestalt values for the unidentified person into the computer. This can be done by either processing a picture of the person immediately before the recognition process, or loading a record of values for this person from the Processed Picture Database. (See menu item #4.)

2) To "identify" the person, select menu item #5. Table B-2 is an example of the output from this process.

For each window, a list is developed for potential candidates. The closeness of the unidentified person's gestalt coordinates to each of the candidate's is expressed as a pseudo-probability. (This probability is also weighted by a "performance factor" for each window (see chapter 4.) In table B-2, the "position" for each person indicates the average gestalt values for that person. "Sigmas Away" for X & Y indicate the number of standard deviations the unknown person's gestalt coordinates are away from the candidate's coordinates (in terms of the candidate's X & Y standard

```
          * * * CTT FACE RECOGNITION SYSTEM * * *
     Date: 11/27/85     Time: 17:33


Filename of picture being recognized = SMALL9.PI



      * * * CANDIDATES FOR WINDOW 1 * * *

X,Y Location of Unidentified Person: 12,43    X Sigma (for Window) =     .67
    Y Sigma (for Window) =     .95
Number of Sigmas Out We're Searching =   3.0
Range of Search:  X Coordinate = 10 to 14.  Y Coordinate = 40 to 46.

ID Number =   16   CAPT JIM HOLTEN          Position = 13,40   Prob =    5.17
        Sigmas Away -- X:    1.58     Sigmas Away -- Y:    1.28
ID Number =   13   CAPT PHIL FITZJARREL     Position = 13,40   Prob =     .06
        Sigmas Away -- X:    2.00     Sigmas Away -- Y:    6.00
ID Number =    3   CAPT MAX HALL            Position = 14,40   Prob =     .30
        Sigmas Away -- X:    4.00     Sigmas Away -- Y:    3.29
ID Number =   10   MR. SWAMI KRISHNASWAMI   Position = 11,44   Prob =    6.66
        Sigmas Away -- X:    1.25     Sigmas Away -- Y:     .74
ID Number =    4   CAPT JERRY GERACE        Position = 13,45   Prob =    2.13
        Sigmas Away -- X:    1.61     Sigmas Away -- Y:    2.94
ID Number =    1   CAPT RON SMALL           Position = 12,46   Prob =    5.76
        Sigmas Away -- X:     .00     Sigmas Away -- Y:    1.81
ID Number =   11   CAPT FRED STIERWALT      Position = 14,46   Prob =     .12
        Sigmas Away -- X:    4.00     Sigmas Away -- Y:    4.28



      * * * CANDIDATES FOR WINDOW 2 * * *

X,Y Location of Unidentified Person: 10,43    X Sigma (for Window) =     .87
    Y Sigma (for Window) =    1.00
Number of Sigmas Out We're Searching =   3.0
Range of Search:  X Coordinate =  7 to 13.  Y Coordinate = 40 to 46.

ID Number =   11   CAPT FRED STIERWALT      Position = 10,40   Prob =    3.29
        Sigmas Away -- X:     .00     Sigmas Away -- Y:    2.70
ID Number =    4   CAPT JERRY GERACE        Position = 11,43   Prob =    6.83
        Sigmas Away -- X:    1.20     Sigmas Away -- Y:     .00
ID Number =   18   CAPT RIC ROUTH           Position = 11,45   Prob =    2.65
        Sigmas Away -- X:     .95     Sigmas Away -- Y:    2.85
ID Number =    1   CAPT RON SMALL           Position = 11,45   Prob =    5.79
        Sigmas Away -- X:    1.17     Sigmas Away -- Y:    1.18
```

Table B-2. Example Output From Recognition Program (Part 1)

```
ID Number =    2   CAPT BOB RUSSEL           Position = 10,46   Prob =    1.86
        Sigmas Away -- X:     .00     Sigmas Away -- Y:    3.44



     * * * CANDIDATES FOR WINDOW 3 * * *

X,Y Location of Unidentified Person:  9,34    X Sigma (for Window) =     .81
    Y Sigma (for Window) =    1.93
Number of Sigmas Out We're Searching =   3.0
Range of Search:  X Coordinate =  7 to 11.  Y Coordinate = 28 to 40.

ID Number =    7   CAPT DAVE HUNSUCK         Position = 11,35   Prob =     .57
        Sigmas Away -- X:    4.00     Sigmas Away -- Y:     .46
ID Number =    1   CAPT RON SMALL            Position = 10,36   Prob =    3.18
        Sigmas Away -- X:     .90     Sigmas Away -- Y:    1.28



     * * * CANDIDATES FOR WINDOW 4 * * *

X,Y Location of Unidentified Person: 12,33    X Sigma (for Window) =    1.04
    Y Sigma (for Window) =    1.49
Number of Sigmas Out We're Searching =   3.0
Range of Search:  X Coordinate =  9 to 15.  Y Coordinate = 29 to 37.

ID Number =    1   CAPT RON SMALL            Position = 14,31   Prob =    2.44
        Sigmas Away -- X:    1.85     Sigmas Away -- Y:     .98



     * * * CANDIDATES FOR WINDOW 5 * * *

X,Y Location of Unidentified Person: 16,33    X Sigma (for Window) =    1.92
    Y Sigma (for Window) =    2.30
Number of Sigmas Out We're Searching =   3.0
Range of Search:  X Coordinate = 10 to 22.  Y Coordinate = 26 to 40.

ID Number =   13   CAPT PHIL FITZJARREL      Position = 19,26   Prob =     .00
        Sigmas Away -- X:    2.04     Sigmas Away -- Y:    8.53
ID Number =    8   CAPT MARK CLIFFORD        Position = 17,27   Prob =    1.33
        Sigmas Away -- X:     .29     Sigmas Away -- Y:    2.64
ID Number =    9   DR. WOODROW W. BLEDSOE    Position = 20,27   Prob =     .03
        Sigmas Away -- X:    4.93     Sigmas Away -- Y:    3.68
ID Number =   10   MR. SWAMI KRISHNASWAMI    Position = 18,29   Prob =    2.81
        Sigmas Away -- X:     .70     Sigmas Away -- Y:     .77
```

Table B-2. Example Output From Recognition Program (Part 2)

```
ID Number =   12   CAPT MIKE HUNSUCKER        Position = 21,29   Prob =     .49
        Sigmas Away -- X:     3.16     Sigmas Away -- Y:     2.24
ID Number =   18   CAPT RIC ROUTH             Position = 20,30   Prob =    1.60
        Sigmas Away -- X:     1.52     Sigmas Away -- Y:     1.81
ID Number =    4   CAPT JERRY GERACE          Position = 16,31   Prob =    3.05
        Sigmas Away -- X:      .00     Sigmas Away -- Y:      .65
ID Number =   17   DR. BILL CZELEN            Position = 17,31   Prob =    1.84
        Sigmas Away -- X:     2.00     Sigmas Away -- Y:      .69
ID Number =   16   CAPT JIM HOLTEN            Position = 20,31   Prob =    1.93
        Sigmas Away -- X:     1.91     Sigmas Away -- Y:      .66
ID Number =   20   MRS. EDIE ROUTH            Position = 21,31   Prob =    2.13
        Sigmas Away -- X:     1.51     Sigmas Away -- Y:     1.00
ID Number =    7   CAPT DAVE HUNSUCK          Position = 16,32   Prob =    3.16
        Sigmas Away -- X:      .00     Sigmas Away -- Y:      .39
ID Number =    3   CAPT MAX HALL              Position = 17,32   Prob =    3.08
        Sigmas Away -- X:      .44     Sigmas Away -- Y:      .40
ID Number =    5   CAPT TOM GRIFFIN           Position = 11,37   Prob =    2.01
        Sigmas Away -- X:     1.66     Sigmas Away -- Y:     1.00
ID Number =    2   CAPT BOB RUSSEL            Position = 11,37   Prob =     .31
        Sigmas Away -- X:     3.59     Sigmas Away -- Y:     2.40
ID Number =   14   CAPT DAVID KING            Position = 10,38   Prob =     .33
        Sigmas Away -- X:     3.68     Sigmas Away -- Y:     2.18
ID Number =    6   DR TERRY SKELTON           Position = 12,39   Prob =     .50
        Sigmas Away -- X:     2.87     Sigmas Away -- Y:     2.59
ID Number =    1   CAPT RON SMALL             Position = 14,40   Prob =    1.58
        Sigmas Away -- X:      .78     Sigmas Away -- Y:     2.25


   * * * CANDIDATES FOR WINDOW 6 * * *

X,Y Location of Unidentified Person: 20,32    X Sigma (for Window) =    1.50
    Y Sigma (for Window) =     1.01
Number of Sigmas Out We're Searching =    3.0
Range of Search:  X Coordinate = 16 to 24.  Y Coordinate = 29 to 35.

ID Number =   11   CAPT FRED STIERWALT        Position = 18,30   Prob =     .85
        Sigmas Away -- X:     1.55     Sigmas Away -- Y:     4.00
ID Number =    4   CAPT JERRY GERACE          Position = 19,33   Prob =    7.20
        Sigmas Away -- X:      .95     Sigmas Away -- Y:      .65
ID Number =    1   CAPT RON SMALL             Position = 21,35   Prob =    3.09
        Sigmas Away -- X:      .65     Sigmas Away -- Y:     2.77


  * * * COMPUTER'S CHOICE(S) FOR WHO THIS IS * * *


ID Number =    1   CAPT RON SMALL                  Value =  21.85   % =   .26
```

Table B-2. Example Output From Recognition Program (Part 3)

ID Number =	4	CAPT JERRY GERACE	Value = 19.21	% =	.23
ID Number =	10	MR. SWAMI KRISHNASWAMI	Value = 9.47	% =	.11
ID Number =	16	CAPT JIM HOLTEN	Value = 7.10	% =	.08
ID Number =	11	CAPT FRED STIERWALT	Value = 4.26	% =	.05
ID Number =	18	CAPT RIC ROUTH	Value = 4.25	% =	.05
ID Number =	7	CAPT DAVE HUNSUCK	Value = 3.73	% =	.04
ID Number =	3	CAPT MAX HALL	Value = 3.38	% =	.04
ID Number =	2	CAPT BOB RUSSEL	Value = 2.18	% =	.03
ID Number =	20	MRS. EDIE ROUTH	Value = 2.13	% =	.03
ID Number =	5	CAPT TOM GRIFFIN	Value = 2.01	% =	.02
ID Number =	17	DR. BILL CZELEN	Value = 1.84	% =	.02
ID Number =	8	CAPT MARK CLIFFORD	Value = 1.33	% =	.02
ID Number =	6	DR TERRY SKELTON	Value = .50	% =	.01
ID Number =	12	CAPT MIKE HUNSUCKER	Value = .49	% =	.01
ID Number =	14	CAPT DAVID KING	Value = .33	% =	.00
ID Number =	13	CAPT PHIL FITZJARREL	Value = .06	% =	.00
ID Number =	9	DR. WOODROW W. BLEDSOE	Value = .03	% =	.00

Table B-2. Example Output From Recognition Program (Part 4)

deviations.) See chapter 4 for a detailed explanation.

The computer's choice is made by adding the probabilities for each candidate for each window, and then sorting them from highest to lowest. The candidate with the highest value wins. The confidence that the highest person on the ist is the correct one can be judged by how far ahead the top person's value is from the contenders. If the correct person is not first, then he or she is usually one of the top candidates.

The following describes the process involved in loading a record from the database for recognition, and recognizing a person.

Load a Record from Database for Recognition.

```
Enter number of record you would like to load:  42
Record Number: 42   ID Number: 3   Dr. Matthew Kabrisky
Is this the record you want? (1=Yes,2=No): 1
```

At this point, the record for Dr. Kabrisky is residing on the computer, as if his picture had just been processed.

Identify a Person

```
* * * CTT Face Recognition System * * *
```

Do you want to use window performance factors?
(1=Yes,2=No): 1

(Indicate Yes for the normal recognition process.)

Do you want the results printed? (1=Yes,2=No): 1

How do you want to enter data for the face to be recognized?

1 - Use data last loaded on computer
2 - Enter Gestalt values manually

Choice: 1

Retrieving Data...

An example of the resulting output is shown in Table B-2.

Database Maintenance

The program MAIN includes some database maintenance and testing functions under the Data Base Maintenance Menu.

* * * Data Base Maintenance Menu * * *

1 - Examine Contents of Recognition Database
2 - Add a Record to Recognition Database
3 - Make a New Recognition Database

4 - Change Data Elements in Processed Picture Database

5 - Mark Records for Deletion

9 - Delete Records Marked for Deletion

-1 - Return to Main Menu

Choice:

Quit

When the user selects "Quit", the program MAIN terminates, and the directory changes back to ERUSSEL.

Appendix C -- Calibration

In order to find a proper equipment setup and calibration procesure, the following was investigated:

1) Determination of proper camera settings
2) Insuring consistent lighting conditions

As a result, an equipment set-up and a calibration procedure were developed which seem to minimize the error due to camera settings and insure consistent lighting conditions.

Determining proper camera-to-image distance. The goal of this investigation was to find the range of zoom adjustments which made the least effect on average pixel value, and then pick the camera-to-image distance which exploited this zoom range. The size of a target image was first measured by physically moving the camera different distances, with the zoom set to its lowest value (18mm). Then the camera was moved a fixed distance away, and the zoom adjusted to achieve standard image size. For instance, assume a square object showed dimensions of 6" x 6" on the monitor, at a distance of 6 feet from the camera. To find the camera's equivalent zoom distance with the camera 9 feet away from the image, adjust the zoom until the object is once again 6" x 6" on the monitor.

The average pixel value was plotted for equivalent zoom distances of from 4 to 9 feet (see figure C-1.) The range of the curve with least variation was found to be from 6.5 to 9 feet. It was also found that a range of about 6-8 feet away from the subject, with the zoom set to its lowest value

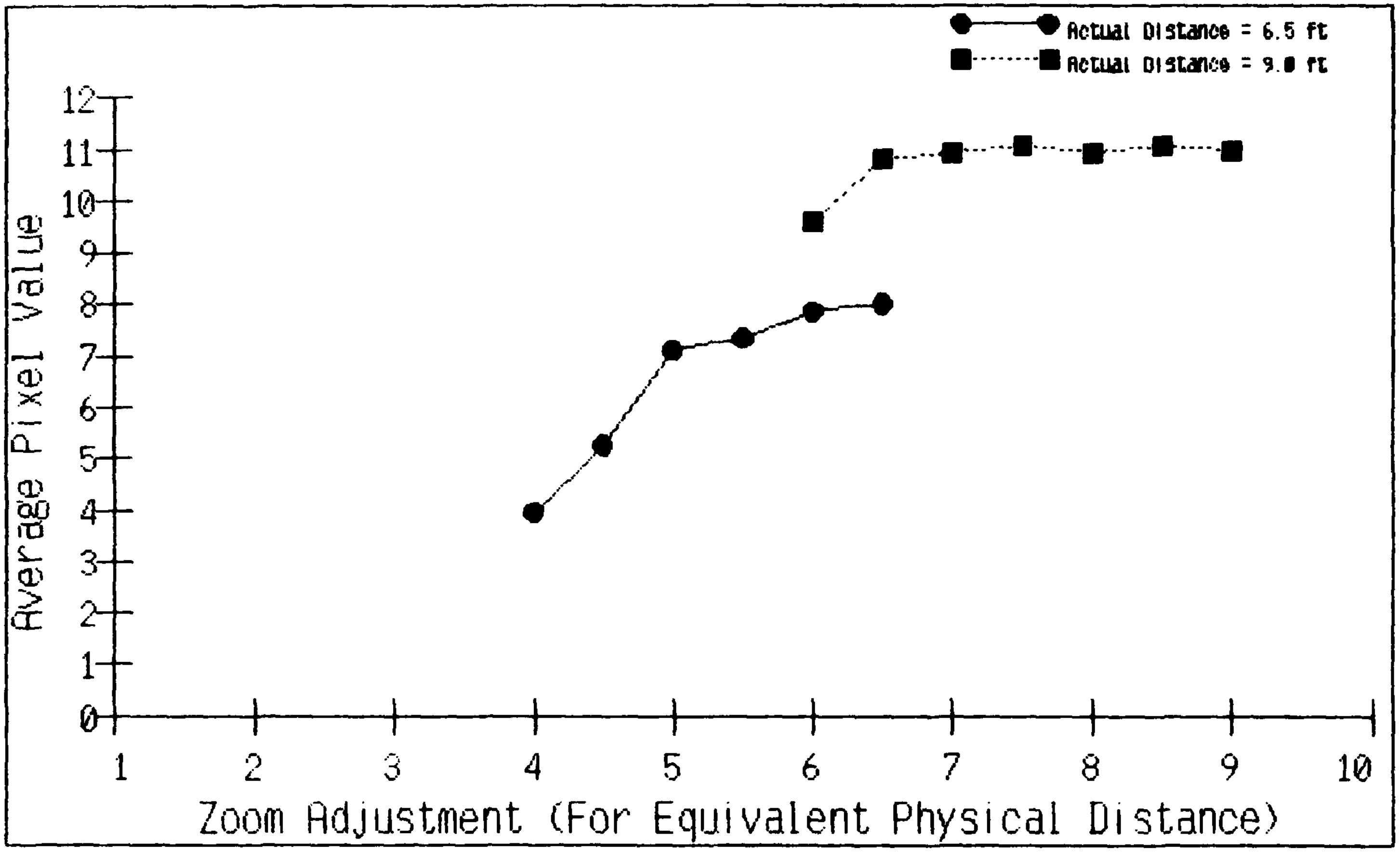


Figure C-1. Zoom Adjustment vs Average Pixel Value

(18mm), allowed most facial images to be small enough to fit within the 64x64 pixel box cursor. As a result of the study, it was determined that a distance of 8 feet from the camera to the target provided an adequate range of zoom adjustments with minimal change in average pixel value.

Determining proper camera warm-up time. The average pixel value of light reflected from an 18% gray card was measured over a period of 3 hours, starting from when the camera was first turned on. Figure C-2 shows a graph of values measured over one hour. The conclusion was that a 10-15 minute warm-up period was necessary to minimize changes in average pixel value over the period of time of the picture-taking process. In addition, a calibration was performed on images before processing, to insure proper warm-up and consistent lighting conditions. This was also done by measuring reflected light from an 18% gray card, as described below:

CALIBRATION PROCEDURES

Test Setup:

FOCUS = 30 feet

F-STOP = F5.6

ZOOM = 18mm

Distance from gray card to camera focal plane = 6.0 feet

Once the above parameters are set up, the average pixel value of the gray card is measured. The result is usually between 9.0 and 10.0 with the camera warmed up in the Signal Processing Lab. If it is not, either the camera probably

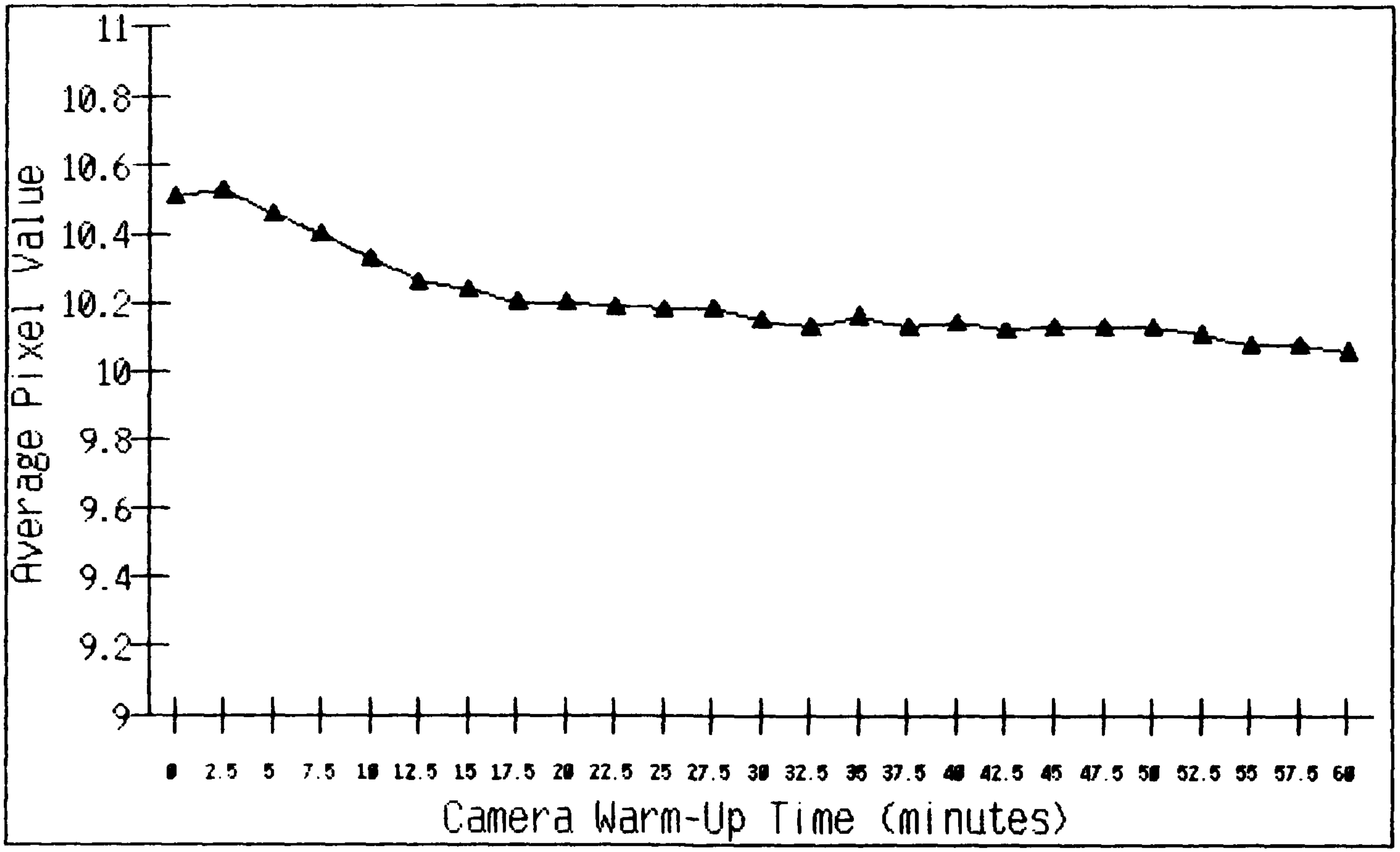


Figure C-2. Camera Warm-Up Time vs Average Pixel Values

needs warming up, or possibly some of the overhead lights have burned out.

(See Appendix B, "User's Guide", for detailed instructions on how to measure the average pixel value on an image.)

Determining effect of focus on average pixel value. To Measure this effect, the zoom was adjusted to "equivalent physical distances," as explained in the previous paragraph. The camera-to-image distance was 9 feet. The average pixel value was measured over a focus range of 4 feet to infinity, with the zoom set to different values (see figure C-3). It was found that the average pixel value varied little over its range with zoom seetings from 7 feet to 9 feet. However, from 6 feet to 7 feet, the average pixel value began to vary from 1-2 pixels over its range.

As mentioned in paragraph 2 above, a range of physical distances (or zoom equivalent) of 6-9 feet was found sufficient for measuring human faces. Because the average pixel value changed too much within the range 6-7 feet, however, it was decided to fix the focus value at 9 feet. This value provided an adequate focus for images 6-9 feet away, while minimizing the change in average pixel value.

F-Stop Setting. F-stop is set to F8.0. The reason for this is explained in Chapter 4 under "Contrast Enhancement."

Summary of Camera Settings.

FOCUS = 9 feet

ZOOM = Variable

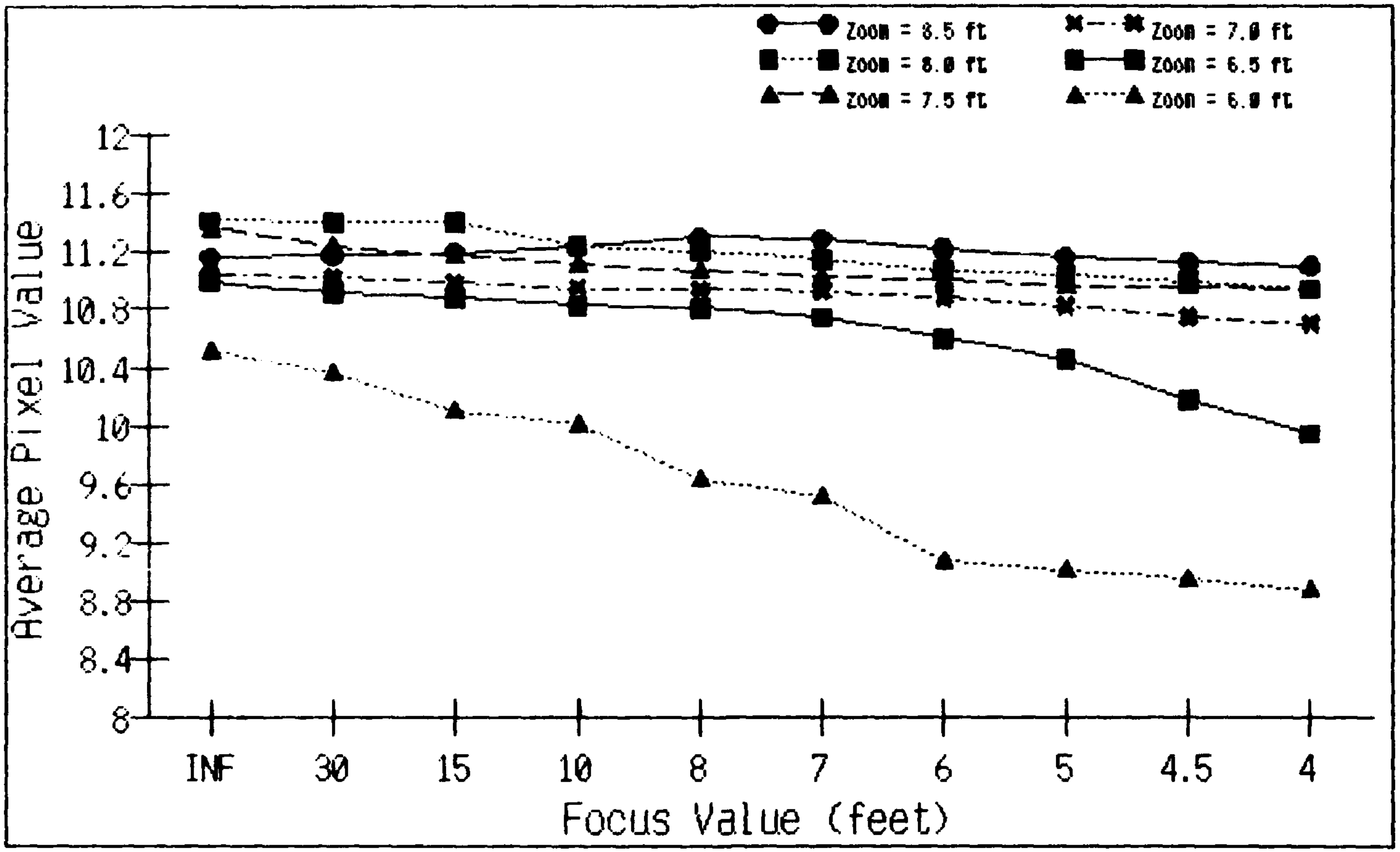


Figure C-3. Focus Value vs Average Pixel Value

F-STOP = 8.0

Distance from image to focal plane of camera = 9 feet (or to center of tripod = 9'6")

The final studio setup is displayed in figure C-4.

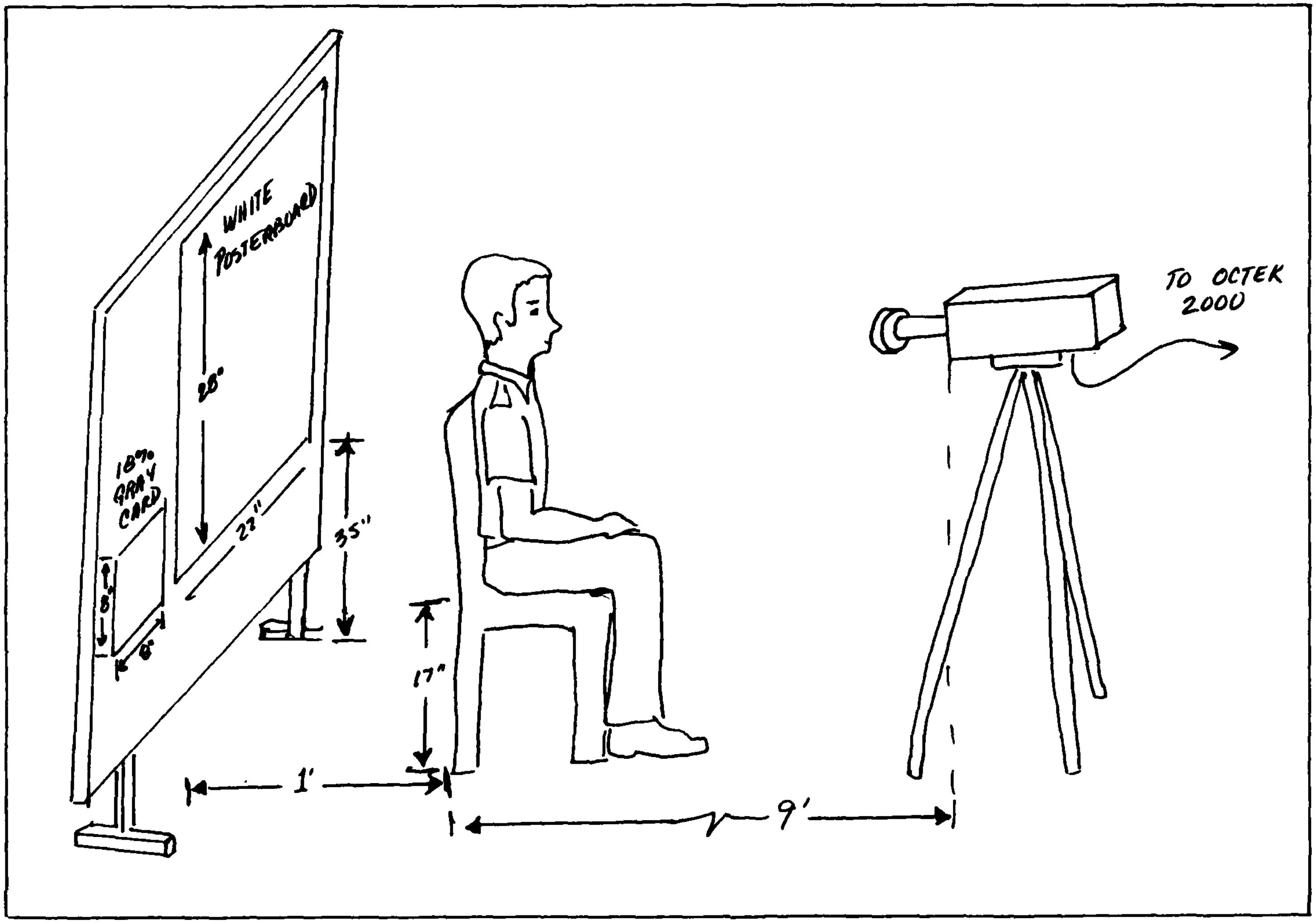


Figure C-4. Studio Setup For Taking A Picture

Appendix D -- File Generation and Communications

This appendix describes the file generation and file communications used in the CTT Face Recognition System. In order to allow communications between the Eclipse computer and the Nova computer, the computers generate files in a shared disk drive. Files are also used to pass data between different programs on the same computer.

The following is a list of the different files and their purposes:

FILE	FIGURE	PURPOSE
NOVASIG1	D-1a, D-1b, D-2	Contains filename of picture file
FSTOP	D-1a, D-1c	Contains FSTOP value of picture file
WINDOWLOC	D-1a,D-1b, D-1c, D-2	Contains locations of features on face
MULTVAL	D-1a,D-1c, D-2	Contains contrast-multiplier value for final contrast-expanded image
NOVASIG1.A to NOVASIG6.A	D-1b, D-1c	Flag files created by the program PROCESS2 on the NOVA to signal that the sub-image file WIND1.PI to WIND6.PI are generated and ready for processing by the ECLIPSE.
WIND1.PI to WIND6.PI	D-1b	Sub-image files generated by PROCESS2, which contain sub-images of faces.
COORDPTS1.B to COORDPTS6.B	D-1b, D-1c, D-2,D-4	Contain gestalt coordinates processed from files WIND1.PI to WIND6.PI.

FILE	FIGURE	PURPOSE
PRNTIMAGE	D-1c	Flag file sent from NOVA to signal ECLIPSE to print the image file "TEMP.VD" on the ECLIPSE line printer.
TEMP.VD	D-1c	Stored screen image of completed picture processing
STOREREC	D-1c	Flag file which (if existing) tells program "TRAIN" on NOVA to load the record just processed into the MAINPICS database.
IDNUM	D-1c	Contains ID number of individual (loaded by user in WRNAME)
IDFILE	D-1c, D-3, D-4	Contains an ID number and name for all individuals in database
MAINPICS	D-1c, D-2, D-3	Contain records of processed pictures
FACEDONE	D-1d	Flag file generated by NOVA program "QUIT", which stops CORTRAN16 on the ECLIPSE
WINDOW1 to WINDOW6	D-3, D-4	Part of the coordinate database which stores a pointer to the lookup table in WINDOW1.LU to WINDOW6.LU
WINDOW1.LU to WINDOW6.LU	D-3, D-4	The lookup table for the database, which contains the ID number and standard deviation data for individuals
WINDOW1.SP to WINDOW6.SP	D-3, D-4	Part of the coordinate database holding the location of the next free location in the lookup table.
SIGMAS	D-3, D-4	Contains average of all X & Y standard deviations for each window

WFACTOR	D-3, D-4	Contains window performance factors for each window

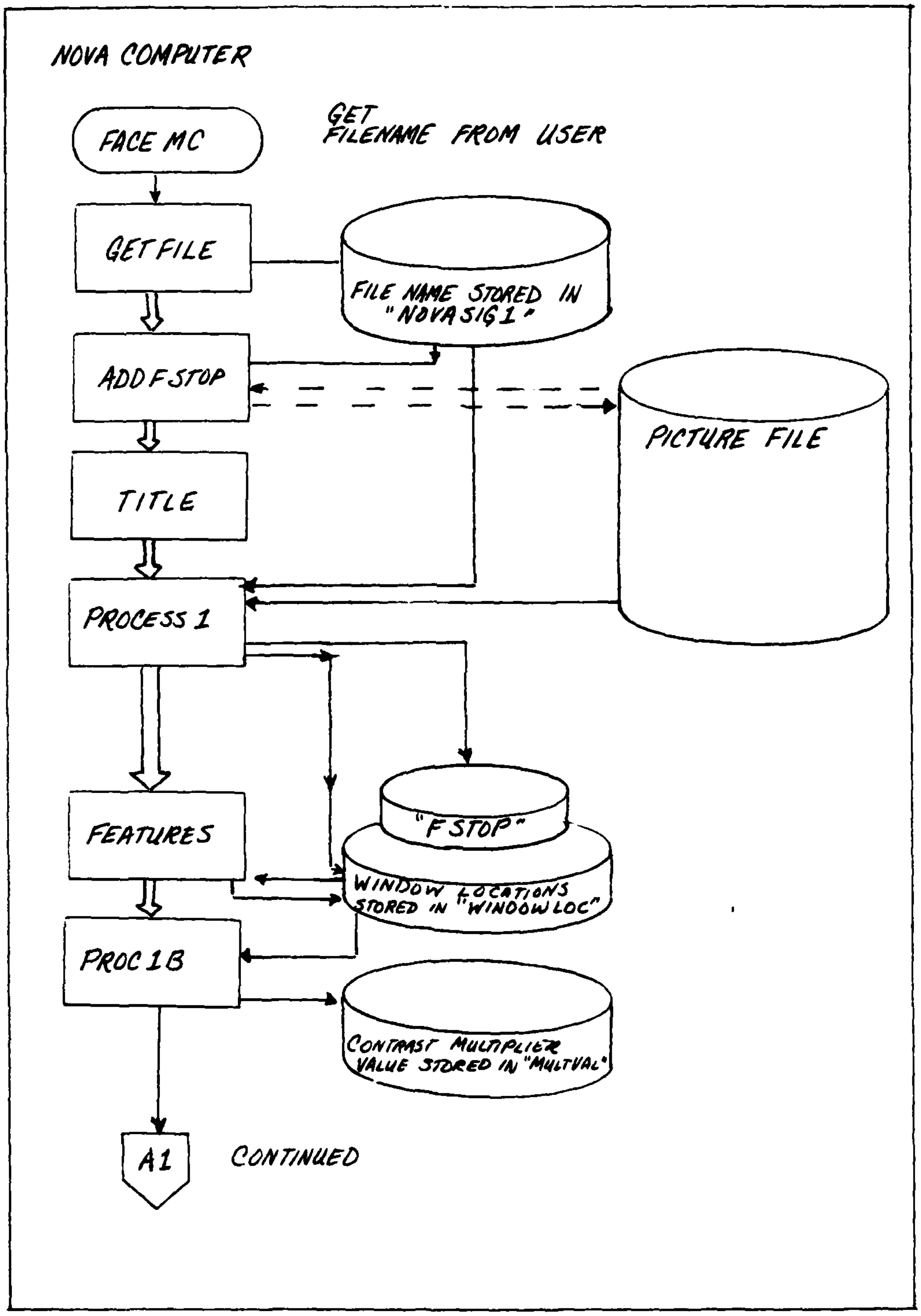


Figure D-1a. System Program Flow and File Use (Part 1)

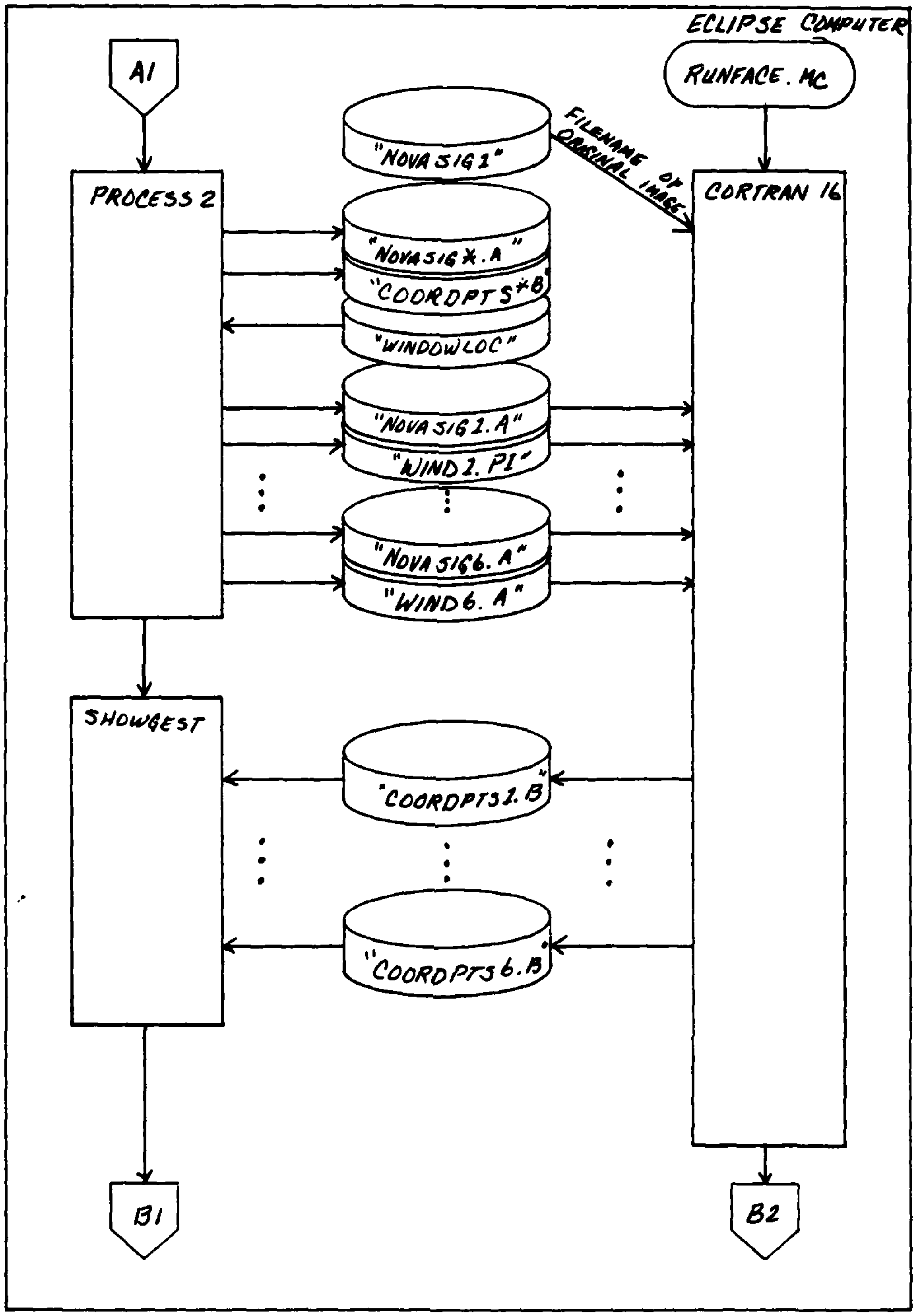


Figure D-1b. System Program Flow and File Use (Part 2)

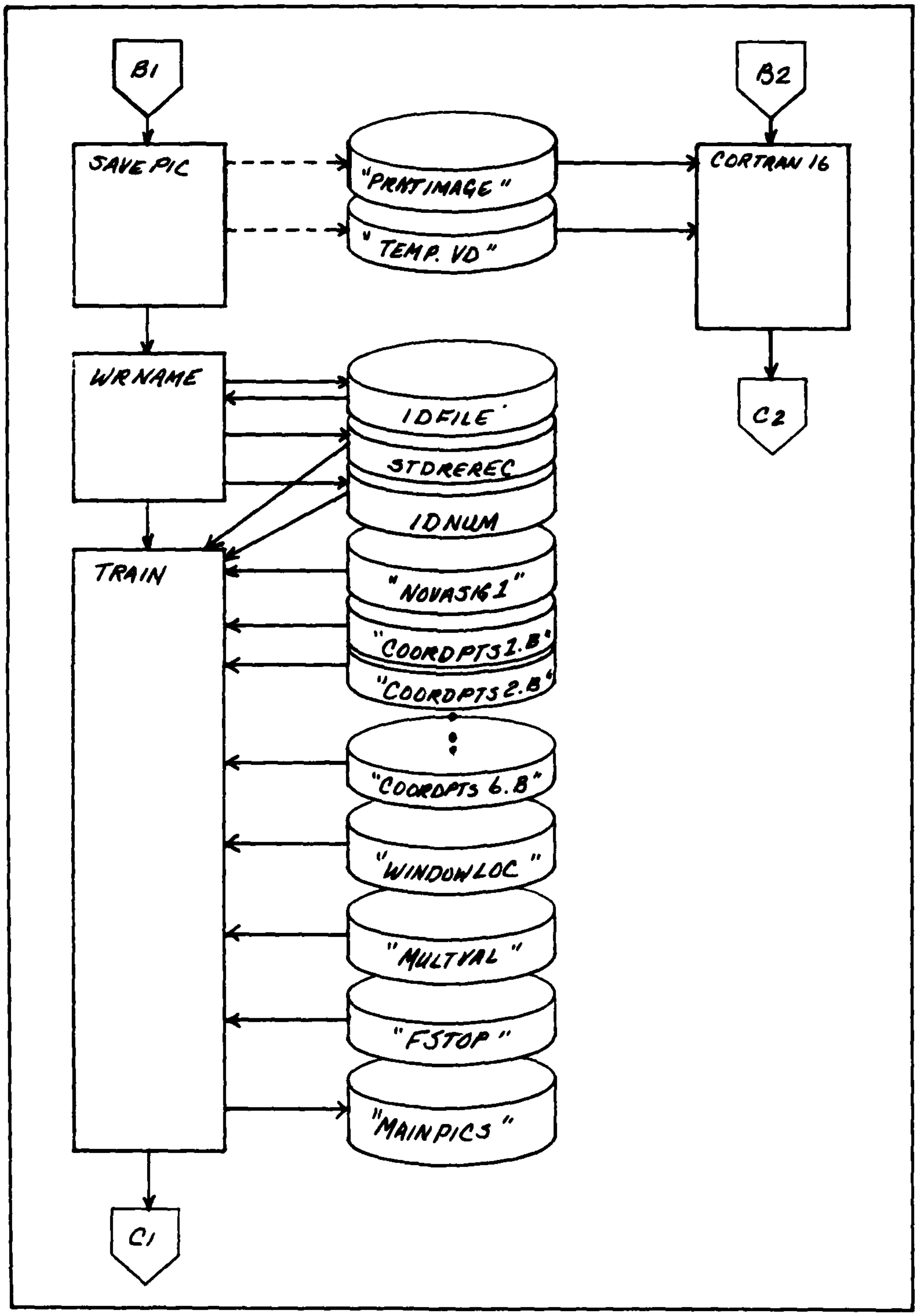


Figure D-1c. System Program Flow and File Use (Part 3)

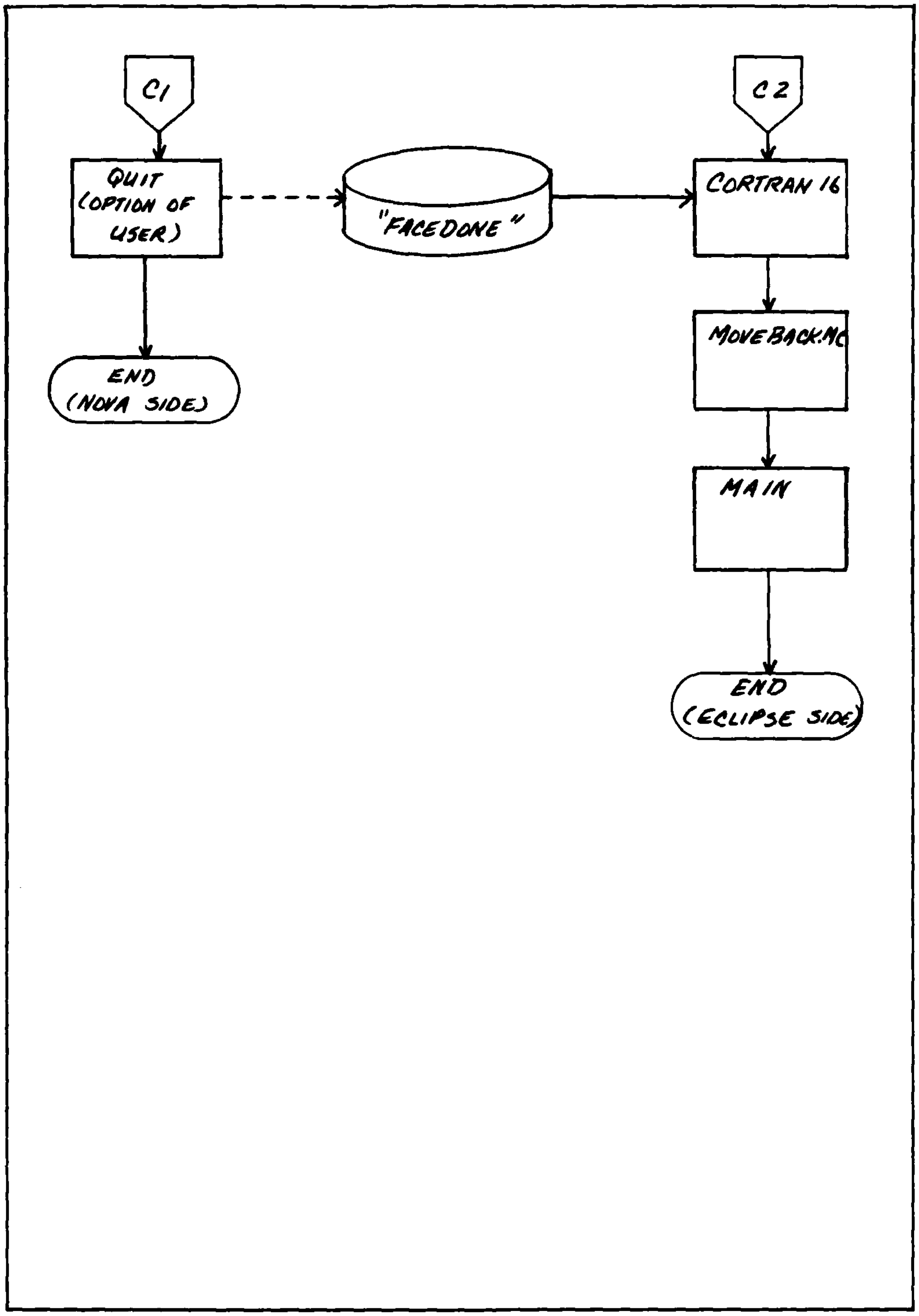


Figure D-1d. System Program Flow and File Use (Part 4)

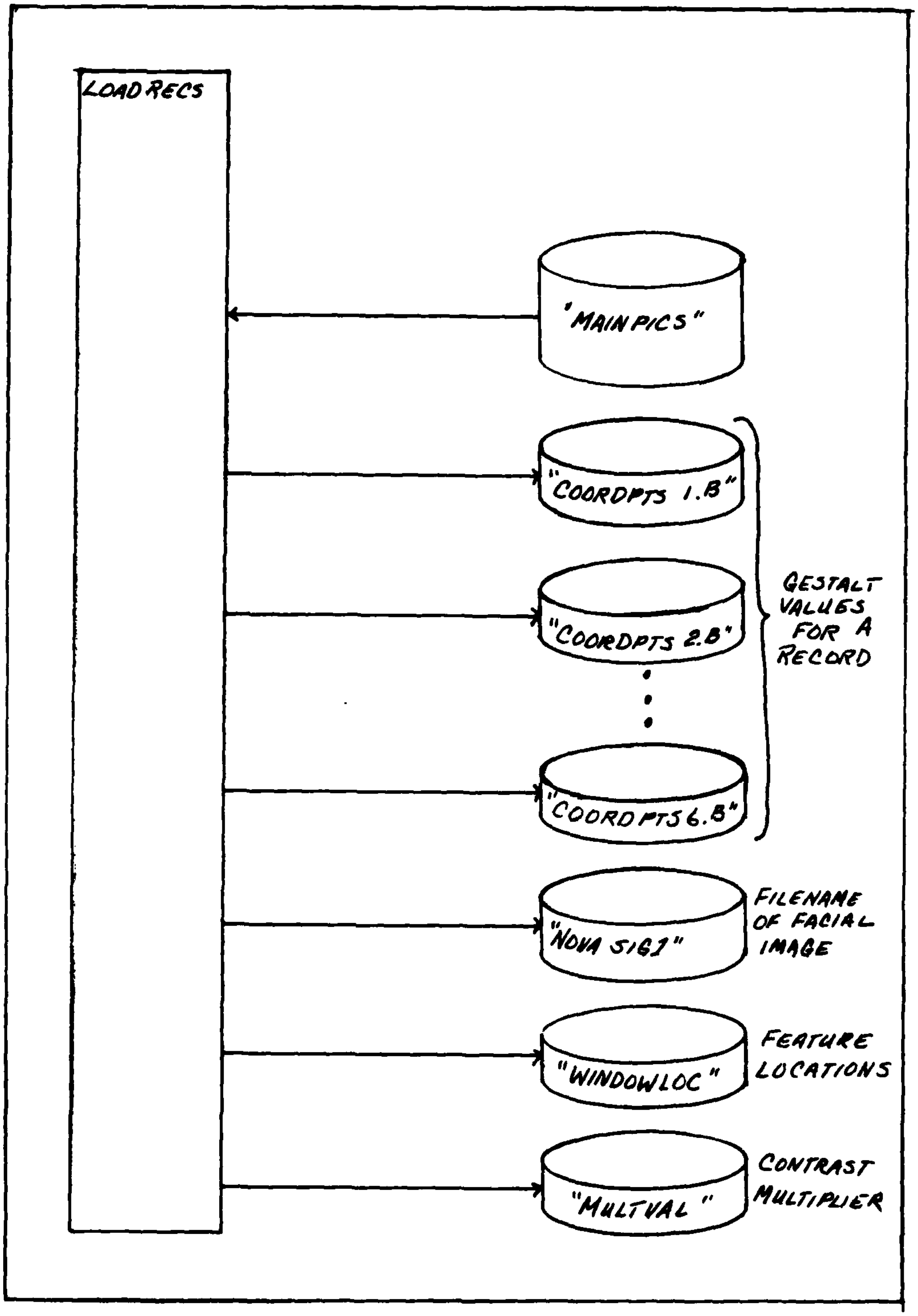


Figure D-2. Disk File Interface For Subroutine "LOADRECS"

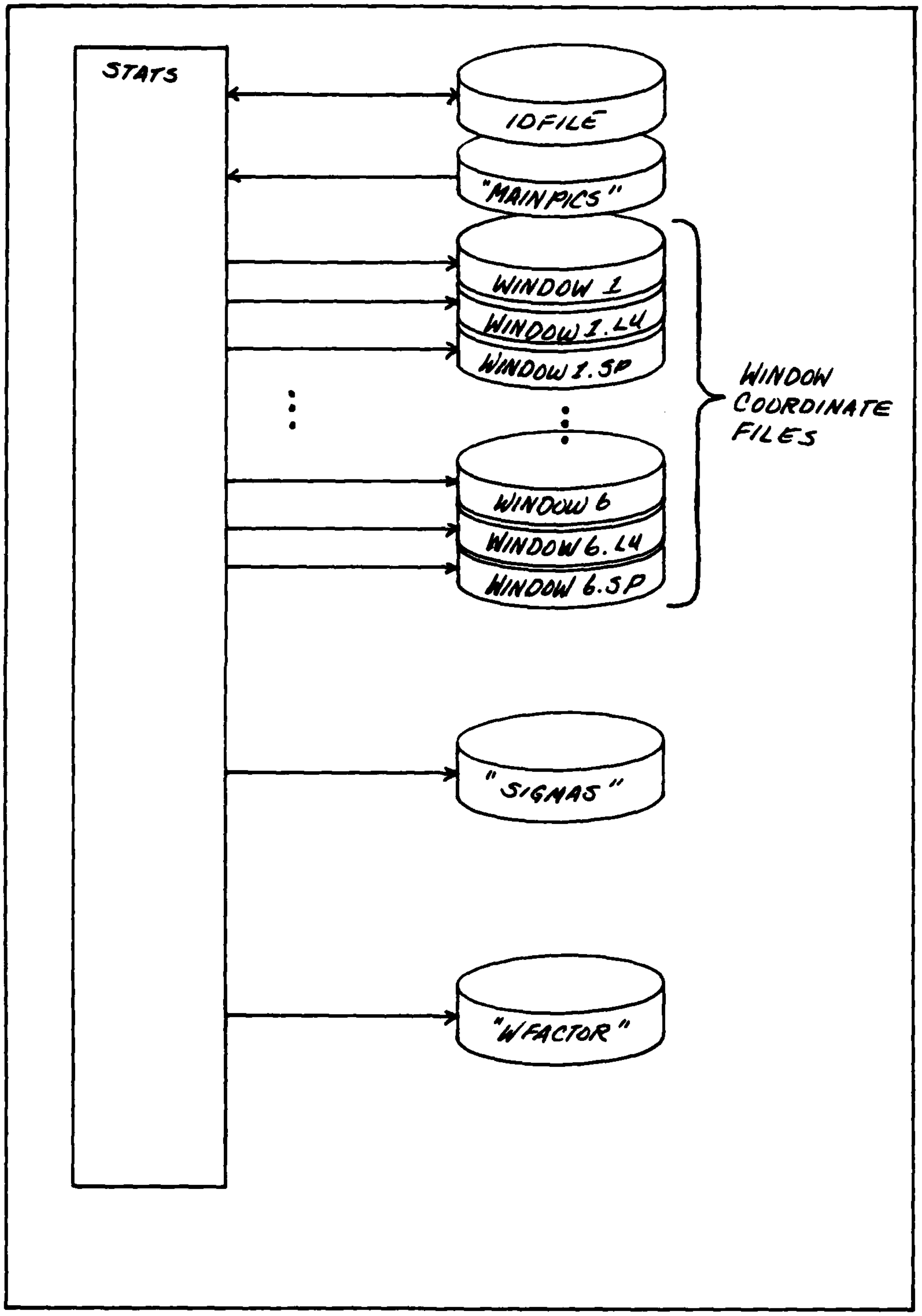


Figure D-3. Disk File Interface For Subroutine "STATS"

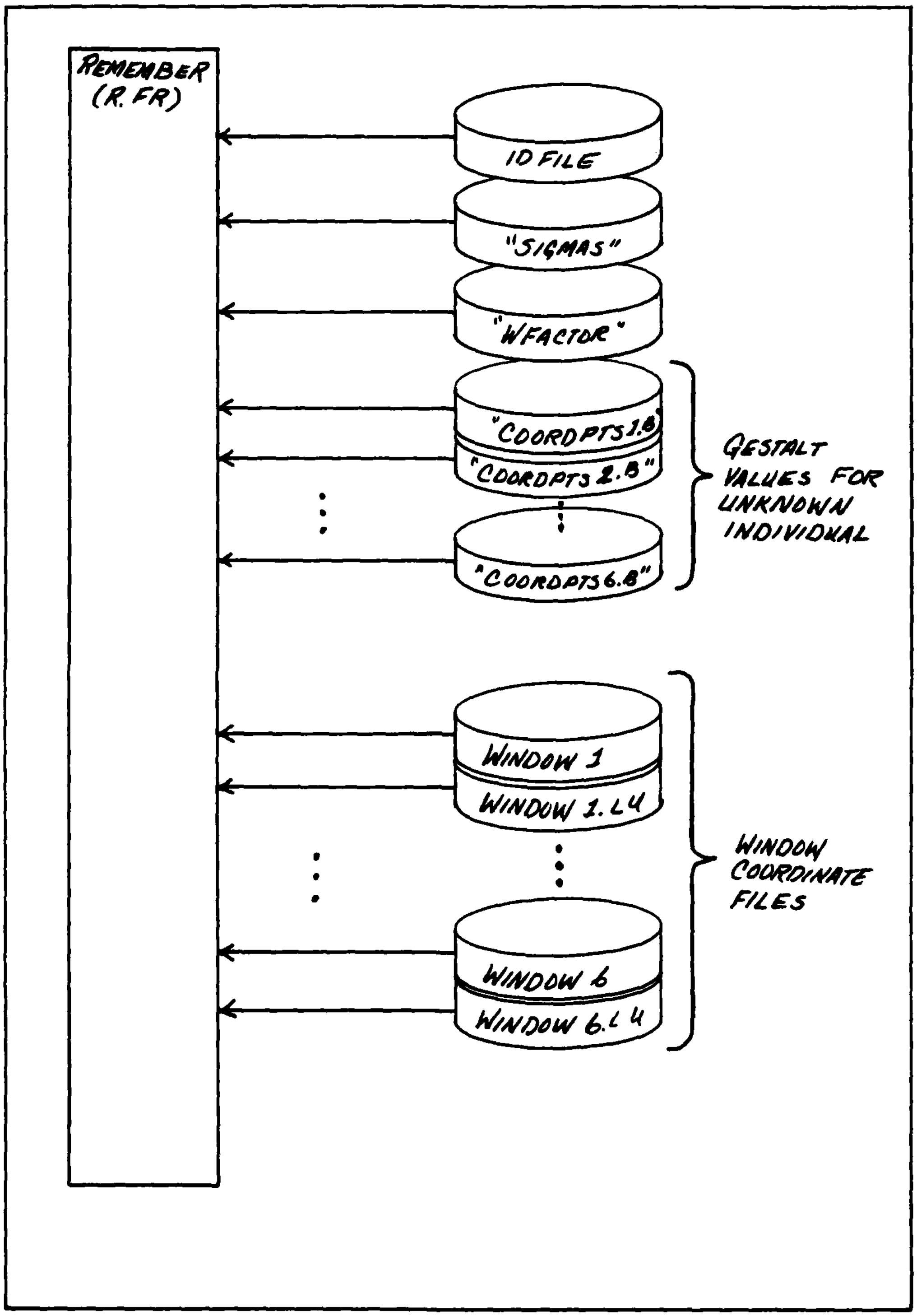


Figure D-4. Disk File Interface For Subroutine "REMEMBER"

Appendix E -- How to Define a Sub-Image on the Face

Defining which sub-images will be processed takes place in PROCESS2. A sub-image can be built from any combination of the feature locations stored in WINDOWLOC. The feature-finding subroutine FINDW initially stores more feature locations than are actually displayed and used by the system. (These other locations may not be accurate, either, as the user cannot update them using the program FEATURES.) The locations actually used are listed below in Table E-1:

Location in File	Feature
1	Top of Head
3	Eyes Begin
4	Eyes End
5	Top of Nose
8	Center of Mouth
11	Lowest light area on Chin
15	Center of Face (between eyes)
16	Left side of Left Eye
17	Right side of Right Eye
18	Left side of Head
19	Right side of Head

Table E-1. Feature Locations used by System

The horizontal locations (left side, right side, center

of face) indicate the number of colums from the left side of the image. The vertical locations indicate the number of rows down from the top of the image. For example, if element 1 of WINDOWLOC has a value of 3, this means the top of the head begins on the 3rd row from the top of the image.

The subroutine call RDWIN defines the subpart of the face accessed. The parameters of this subroutine are as follows:

```
CALL RDWIN (ICT,IXC,IYC,IPIX,IHORT,IVERT,ILEFT,
     IRIGHT,ITOP,IBOT)
```

The boundaries of the sub-image are defined by ILEFT, IRIGHT, ITOP, and IBOT, where these represent the locations of the left, right, top, and bottom of the image, respectively. To change the boundaries for an image, substitute new feature locations from WINDOWLOC for these variables. IHORT, IVERT, and IPIX are returned by the subroutine. ICT, IXC, and IYC are previously defined and need not be redefined.

If the reader decides to modify the feature locations, he should realize this will invalidate the present data stored in MAINPICS, and new data will have to be calculated. If additional sub-images are desired, then the entire system will have to be modified. This should be straightforward,

though time-consuming.

The system can reasonably be extended to any number of new images using the same principles used for the present six images. Extension to additional windows should provide the system better discrimination ability and, if the theory is right, a closer approximation to the human visual system.

Appendix F -- Structure of Record in Processed Picture Database (MAINPICS)

Positions of data in data block:

1 - ID number of person
2-11 - Filename of picture being processed

* * * Window 1 Data * * *
12 - Window Size
13 - Gestalt value (not adjusted)
14 - Gestalt value (adjusted by X & Y)
15 - Gestalt value (adjusted by max of X & Y)
16 - Amplitude

* * * Window 2 Data * * *
17 - Window Size
18 - Gestalt value (not adjusted)
19 - Gestalt value (adjusted by X & Y)
20 - Gestalt value (adjusted by max of X & Y)
21 - Amplitude

* * * Window 3 Data * * *
22 - Window Size
23 - Gestalt value (not adjusted)
24 - Gestalt value (adjusted by X & Y)
25 - Gestalt value (adjusted by max of X & Y)
26 - Amplitude

* * * Window 4 Data * * *
27 - Window Size
28 - Gestalt value (not adjusted)
29 - Gestalt value (adjusted by X & Y)
30 - Gestalt value (adjusted by max of X & Y)
31 - Amplitude

* * * Window 5 Data * * *
32 - Window Size
33 - Gestalt value (not adjusted)
34 - Gestalt value (adjusted by X & Y)
35 - Gestalt value (adjusted by max of X & Y)
36 - Amplitude

* * * Window 6 Data * * *

37 - Window Size
38 - Gestalt value (not adjusted)
39 - Gestalt value (adjusted by X & Y)
40 - Gestalt value (adjusted by max of X & Y)
41 - Amplitude

42-71 - Window Locations on the Face
75 - Value used to multiply for Contrast Expansion

76 - Flag indicating whether or not this record will be used in training the system.

77 - FSTOP times 100 (e.g., F5.6 =56). 999 = Don't Care.

Time this record was processed.
80 - Month
81 - Day
82 - Year
83 - Hour
84 - Minute
85 - Second

Bibliography

1. Bisson, C. L. Location of some facial features by computer, Palo Alto: Panoramic Research, Aug 1966.

2. Bledsoe, W. W. Man-machine facial recognition: Report on a large scale experiment. Palo Alto: Panoramic Research, August 1, 1966.

3. Burke, P. H., Beard, L. F. H. "Stereophotogrammetry of the face," American Journal of Orthodontics. Vol 53, 1967.

4. Galper, Ruth Ellen; Hochberg, Julian. "Recognition Memory for Photographs of Faces," American Journal of Psychology, Vol 84, No. 3, 1971.

5. Ginsburg, Arthur P. "Is the illusory triangle physical or imaginary?" Nature, Vol 257, No. 5523, Sept 18, 1975.

6. Goldschlager, Leslie M. A Computational Theory of Higher Brain Function. Stanford University: STAN-CS-84-1004, April 1984.

7. Goldstein, A. J.; Harmon, L. D.; Lesk, A. B. "Identification of Human Faces," Proceedings of the IEEE, Vol 59, No. 5, May 1971.

8. Goldstein, Alvin G.; Mackenberg, Edmund J. "Recognition of human faces from isolated facial features: A developmental study," Psychonomic Science, Vol 6, No. 4, 1966.

9. Haith, Marshall M.; Bergman, Terry and Moore, Michael J. "Eye Contact and Face Scanning in Early Infancy", Science, Vol 198, 25 Nov 1977.

10. Harmon, L. D. "Some Aspects of Recognition of Human Faces," Pattern Recognition in Biological and Technical Systems. Springer-Verlag, New York, 1971. Not in circulation.

11. Hubel, David H. "The Visual Cortex of the Brain." Scientific American, Nov 1963.

12. Kabrisky, M., et al. "A Theory of Pattern Perception Based on Human Physiology." Contemporary Problems in Perception. Edited by A.T. Welford and L.H. Houssiadas. Taylor and Francis Ltd: London, pp. 129-147, 1970.

13. Kabrisky, Matthew. A Proposed Model for Visual Information Processing in the Human Brain. University of Illinois Press: Urbana, 1966.

14. Kabrisky, Matthew. Class lecture notes, Air Force Institute of Technology, 1984-1985.

15. Krieg, Wendell J.S. Functional Neuroanatomy. The Blakiston Company, Inc., New York, 1953.

16. Maher, Frank A. "A Correlation of Human and Machine Pattern Discrimination." Proceedings of the IEEE 1970 NAECON Conference, pp 260-264, 1970.

17. Marr, David C. Vision. Freeman: New York, 1982.

18. Netter, Frank H. The Nervous System, The Ciba Collection of Medical Illustrations, Vol I, Ciba, 1975.

19. Restak, Richard M. The Brain. Bantam Books, Inc., New York, New York, 1984.

20. Rich, Elaine. Artificial Intelligence. McGraw-Hill: New York, 1983.

21. Routh, Richard LeRoy. Cortical Thought Theory: A Working Model of the Human Gestalt Mechanism. PhD Dissertation, AFIT/DS/EE/85-1, Air Force Institute of Technology,DTIC Document, July 1985.

22. Routh, Richard LeRoy, Robert L Russel, Jr. and Matthew Kabrisky. Cortical Thought Theory: A New Computing Architecture Based on the Human Brain. To be published. Conference on Aerospace Applications of Artificial Intelligence (AAAI), 1985.

23. Sassouni, Viken. "Palatoprint, Physioprint, and Roentgenographic Cephalometry, as New Methods in Human Identification (Preliminary Report)," Journal of Forensic Sciences, Vol 2, No. 4, Oct 1957.

24. Winston, Patrick Henry. Artificial Intelligence. Addison-Wesley: Reading, Massachusetts, 1977.

VITA

Capt Robert L. Russel Jr. is an information systems engineer in the U.S. Air Force. He received a Bachelor of Science degree in Electrical Engineering from the United States Air Force Academy in 1977, along with a commission in the USAF.

He served as a communications officer, with assignments in the areas of technical evaluation of ground communications systems, military applications of fiber optics communications, and command and control. He entered the School of Engineering, Air Force Institute of Technology, in June 1984.

PII Redacted [redacted]

UNCLASSIFIED
SECURITY CLASSIFICATION OF THIS PAGE

REPORT DOCUMENTATION PAGE

1a. REPORT SECURITY CLASSIFICATION UNCLASSIFIED	1b. RESTRICTIVE MARKINGS N/A
2a. SECURITY CLASSIFICATION AUTHORITY	3. DISTRIBUTION/AVAILABILITY OF REPORT APPROVED FOR PUBLIC RELEASE: DISTRIBUTION UNLIMITED.
2b. DECLASSIFICATION/DOWNGRADING SCHEDULE	
4. PERFORMING ORGANIZATION REPORT NUMBER(S) AFIT/GE/ENG/85D-37	5. MONITORING ORGANIZATION REPORT NUMBER(S)
6a. NAME OF PERFORMING ORGANIZATION School of Engineering Air Force Inst. of Technol. 6b. OFFICE SYMBOL (If applicable) AFIT/ENG	7a. NAME OF MONITORING ORGANIZATION
6c. ADDRESS (City, State and ZIP Code) Wright-Patterson AFB, OH 45433	7b. ADDRESS (City, State and ZIP Code)
8a. NAME OF FUNDING/SPONSORING ORGANIZATION 8b. OFFICE SYMBOL (If applicable)	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER
8c. ADDRESS (City, State and ZIP Code)	10. SOURCE OF FUNDING NOS. PROGRAM ELEMENT NO. / PROJECT NO. / TASK NO. / WORK UNIT NO.
11. TITLE (Include Security Classification) see box 19	

12. PERSONAL AUTHOR(S) Robert L. Russel Jr., Captain, USAF			
13a. TYPE OF REPORT Masters Thesis	13b. TIME COVERED FROM ______ TO ______	14. DATE OF REPORT (Yr., Mo., Day) 85 Dec 04	15. PAGE COUNT 260
16. SUPPLEMENTARY NOTATION			

17. COSATI CODES

FIELD	GROUP	SUB. GR.
06	16	
09	02	

18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)

BRAIN COMPUTERS CEREBRAL CORTEX
BIONICS PSYCHOLOGY ARTIFICIAL INTELLIGENCE
(continued on back)

19. ABSTRACT (Continue on reverse if necessary and identify by block number)

Title: Performance of a Working Face Recognition Machine using Cortical Thought Theory

Thesis Chairman: Matthew Kabrisky, PhD
Professor of Electrical Engineering

Approved for public release: IAW AFR 190-1.
LYNN E. WOLAVER 16 JAN 86
Dean for Research and Professional Development
Air Force Institute of Technology (ATC)
Wright-Patterson AFB OH 45433

20. DISTRIBUTION/AVAILABILITY OF ABSTRACT UNCLASSIFIED/UNLIMITED ☒ SAME AS RPT. ☐ DTIC USERS ☐	21. ABSTRACT SECURITY CLASSIFICATION UNCLASSIFIED
22a. NAME OF RESPONSIBLE INDIVIDUAL Matthew Kabrisky, Professor, GS-15	22b. TELEPHONE NUMBER (Include Area Code) 513-255-5276 22c. OFFICE SYMBOL AFIT/ENG

DD FORM 1473, 83 APR EDITION OF 1 JAN 73 IS OBSOLETE.

UNCLASSIFIED
SECURITY CLASSIFICATION OF THIS PAGE

ABSTRACT:

A face recognition system was developed. based on the principles of Cortical Thought Theory (CTT), recently proposed by Dr. Richard Routh as his doctoral dissertation at the Air Force Institute of Technology (Sep 85.) Routh tested the CTT architecture successfully for speech processing. In order to evaluate this architecture as a generic sensory information processing model, CTT was tested for visual processing, specifically for the difficult task of human face recognition.

The CTT gestalt transformation maps a 2-dimensional image into a 2-D coordinate point. The present system extracts six sub-images from a contrast-expanded image, calculates the 2-D gestalt coordinates, and stores the information in a database. Statistics are then calculated on at least five prototypes processed for each person. Overall performance of different sub-windows on a face are also determined. An "unidentified" person is recognized by calculating the six gestalt feature vectors, and then finding the closest match to previously stored data. The computer generates an ordered list by closeness of match. Performance testing of the system yielded a reliability of 90% for a database of 20 people.

The system exhibits many characteristics of human recognition. The following are the significant results of his research:

1) Provides a possible explanation of why the primate visual system splits images vertically before displaying them on separate right and left primary visual cortexes.

2) Provides a plausible explanation of why humans experience difficulty in recognizing negative images.

3) Faces which look similar to humans map close together in CTT space, and faces which look quite differentto humans map far apart in CTT space.

4) Partial face images which seem to give the highest recognition performance in human psychological experiments give the highest performance in the CTT model.

5) The system is reasonably consistent with the human physiology as it is presently understood.

The performance of he face recognition system strongly suggests CTT's general applicability to vision, and increases its credibility as a general model of human sensory information processing.

Continued from block 18 - Subject Terms:

COMPUTER ARCHITECTURE
SPEECH RECOGNITION
IMAGE PROCESSING
PERCEPTION (psycholgy)
LEARNING